Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 167 (2017)

More on bounded independence plus noise: Pseudorandom generators for read-once polynomials

Chin Ho Lee Emanuele Viola

April 5, 2018

Abstract

We construct pseudorandom generators with improved seed length for several classes of tests. First we consider the class of read-once polynomials over GF(2) in m variables. For error ε we obtain seed length $\tilde{O}(\log(m/\varepsilon))\log(1/\varepsilon)$, where $\tilde{O}$ hides lower-order terms. This is optimal up to the factor $\tilde{O}(\log(1/\varepsilon))$. The previous best seed length was polylogarithmic in m and $1/\varepsilon$.

Second we consider product tests $f\colon \{0,1\}^m \to \mathbb{C}_{\le 1}$. These tests are the product of k functions $f_i\colon \{0,1\}^n \to \mathbb{C}_{\le 1}$, where the inputs of the f_i are disjoint subsets of the m variables and $\mathbb{C}_{\le 1}$ is the complex unit disk. Here we obtain seed length $n \cdot \operatorname{poly}\log(m/\varepsilon)$. This implies better generators for other classes of tests. If moreover the f_i have outputs independent of n and k (e.g., $\{-1,1\}$) then we obtain seed length $\tilde{O}(n + \log(k/\varepsilon))\log(1/\varepsilon)$. This is again optimal up to the factor $\tilde{O}(\log 1/\varepsilon)$, while the previous best seed length was $\ge \sqrt{k}$.

A main component of our proofs is showing that these classes of tests are fooled by almost d-wise independent distributions perturbed with noise.

ISSN 1433-8092

1 Introduction

The construction of unconditional pseudorandom generators (PRGs) that fool restricted classes of tests is a fundamental research direction that has found disparate applications. In this work we obtain new generators for several classes of tests. We start with the simplest.

Fooling read-once polynomials. Pseudorandom generators for *polynomials* have been studied since at least the 1993 work by Luby, Veličković, and Wigderson [LVW93], who gave a generator for GF(2) polynomials of size s with error ε and seed length $2^{O(\sqrt{s/\varepsilon})}$. See [Vio07] for an alternative proof. Servedio and Tan [ST18] recently improved the seed length to $2^{O(\sqrt{\log s})} \cdot \log(1/\varepsilon)$, and any significant improvement on the seed length would require breakthrough progress on circuit lower bounds. For low-degree polynomials, better generators are known [BV10, Lov09, Vio09]. In this work we consider *read-once* polynomials on m variables, which are a sum of monomials on disjoint variables. For this class, a generator with seed length polylogarithmic in m and $1/\varepsilon$ is given in [GLS12] and it applies more generally to read-once ACC^0. We obtain a seed length which is optimal up to a factor of $\tilde{O}(\log 1/\varepsilon)$, where $\tilde{O}$ hides factors $\log\log(m/\varepsilon)$. In particular, when ε is not too small, our generator has seed length optimal up to $\operatorname{poly}\log\log m$.

Theorem 1. *There exists an explicit generator $G\colon \{0,1\}^\ell \to \{0,1\}^m$ that fools any read-once GF(2) polynomial with error ε and seed length $\tilde{O}(\log(m/\varepsilon))\log(1/\varepsilon)$.*

A specific motivation for studying read-once polynomials comes from derandomizing space-bounded algorithms, a major line of research in pseudorandomness whose leading goal is proving RL = L. Despite a lot of effort, for general space-bounded algorithms there has been no improvement over the seed length $\geq \log^2 m$ since the classic 1992 paper by Nisan [Nis92]. In fact, no improvement is known even under the restriction that the algorithm uses *constant-space.* Read-once polynomials can be implemented by constant-space algorithms, and were specifically pointed out by several researchers as a bottleneck for progress on space-bounded algorithms, see for example this survey talk by Trevisan [Tre10]. Thus our work can be seen as progress towards derandomizing small-space algorithms. We note that the concurrent work of Chattopadhyay, Hatami, Reingold and Tal [CHRT18] gives a generator for space-bounded algorithms which implies a generator for polynomials with seed length $\tilde{O}(\log^3 m)\log^2(m/\varepsilon)$.

Theorem 1 also holds for polynomials modulo M for any fixed M; and in fact we obtain it as an easy corollary of a more general generator.

Fooling products. We consider tests on m bits that can be written as the product of k bounded functions on disjoint inputs of n bits. Such tests generalize the well-studied *combinatorial rectangles* [AKS87, Nis92, NZ96, INW94, EGL$^+$98, ASWZ96, Lu02, Vio14, GMR$^+$12, GY14] as well as other classes of tests, see [GKM15]. They were introduced in the latter paper by Gopalan, Kane, and Meka who call them *Fourier shapes.* However, in their definition the partition of the m-bit input into the k n-bit inputs to the functions is fixed and known to the generator. Following a recent push for breaking the mold of "fixed-order" tests, we consider such tests under arbitrary order. We call them *product tests* and define them formally next.

Definition 2 (Product tests)**.** *A function $f\colon \{0,1\}^m \to \mathbb{C}_{\le 1}$ is a* product test *with k functions of input length n if there exist k disjoint subsets $I_1, I_2, \dots, I_k \subseteq \{1, 2, \dots, m\}$ of size $\le n$ such that $f(x) = \prod_{i \le k} f_i(x_{I_i})$ for some functions f_i with range in $\mathbb{C}_{\le 1}$. Here $\mathbb{C}_{\le 1}$ is the complex unit disk $\{z \in \mathbb{C} : |z| \le 1\}$, and x_{I_i} are the $|I_i|$ bits of x indexed by I_i.*

Handling arbitrary order is significantly more challenging, because the classical space-bounded generators such as Nisan's [Nis92] only work in fixed order [Tzu09, BPW11]. Our previous work with Haramaty [HLV17] gave the first generators, but in it the dependency on k is poor: the seed length is always $\ge \sqrt{k}$. In this work we improve the dependency on k exponentially, though the results in [HLV17] are still unsurpassed when k is very small, e.g. $k = O(1)$. We actually obtain two incomparable generators.

Theorem 3. *There exists an explicit generator $G\colon \{0,1\}^\ell \to \{0,1\}^m$ that fools any product test with k functions of input length n with error ε and seed length $\tilde{O}(n+\log k)\log(1/\varepsilon)\log(k/\varepsilon) = n\log^{O(1)}(m/\varepsilon)$.*

By the reductions in [GKM15], we also obtain generators that fool variants of product tests where the outputs of the f_i are not simply multiplied but combined in other ways. These variants include generalized halfspaces [GOWZ10] and combinatorial shapes [GMRZ13, De15], extended to arbitrary order. For those we obtain seed length $n^2\log^{O(1)}(m/\varepsilon)$, whereas the previous best was $\ge n\sqrt{k}$ [HLV17]. As this application amounts to plugging the above theorem in previous reductions, we don't discuss it further in this paper and instead refer the reader to Section 6 in [HLV17].

We then give another generator whose seed length is optimal up to a factor $\tilde{O}(\log 1/\varepsilon)$, just like Theorem 1. However, for this we need the functions f_i in the definition of product tests to take constant values. This condition is satisfied by Boolean and most natural functions. For simplicity one can think of the f_i having outputs $\{-1, 1\}$.

Definition 4 (Nice product tests)**.** *A product test as in Definition 2 is* nice *if each function f_i outputs constant values.*

Formally, one should talk about a nice *family* of product tests; but for simplicity we'll just say "nice product test."

Theorem 5. *There exists an explicit generator $G\colon \{0,1\}^\ell \to \{0,1\}^m$ that fools any nice product test with k functions of input length n with error ε and seed length $\tilde{O}(n+\log(k/\varepsilon))\log(1/\varepsilon)$.*

This is the result from which the generator for polynomials in Theorem 1 follows easily.

Bounded independence plus noise. The framework in which we develop these generators was first laid out by Ajtai and Wigderson in their pioneering work [AW89] of constructing generators for AC^0 with polynomial seed length. The framework seems to have been forgotten for a while, possibly due to the spectacular successes by Nisan who gave better and arguably simpler generators [Nis91, Nis92]. It has been recently revived in a series of papers starting with the impressive work by Gopalan, Meka, Reingold, Trevisan, and Vadhan [GMR$^+$12], who use it to obtain a generator for read-once CNF on m bits with error ε and seed length $\tilde{O}(\log(m/\varepsilon))$. This significantly improves on the previously available seed length of $O(\log m)\log(1/\varepsilon)$ when ε is small.

The Ajtai–Wigderson framework goes by showing that the test is fooled by a distribution with limited independence [NN93], *if we perturb it with noise.* (Previous papers use the equivalent language of *restrictions*, we instead follow [HLV17].) Then the high-level idea is to recurse on the noise. This has to be coupled with a separate, sometimes technical argument showing that each recursion simplifies the test, which we address later. Thus our goal is to understand if bounded independence plus noise fools product tests. For our application, it would be convenient to view the distribution as $D + T \wedge U$, the bit-wise XOR of D and $T \wedge U$, where $\wedge$ is bit-wise AND and $T \wedge U$ is a noise vector: if a bit chosen by T is 1 then we set it to uniform. For the application it is important that T is selected pseudorandomly, though the result is interesting even if T is uniform in $\{0,1\}^m$. We now state the result in [HLV17] after defining almost bounded independence.

Definition 6 ((δ, d)-closeness)**.** *The random variables $X_1, \ldots, X_m$ are (δ, b)-*close *to $Y_1, \ldots, Y_m$ if for every $i_1, \ldots, i_b \in \{1, 2, \ldots, m\}$ the b-tuples $(X_{i_1}, \ldots, X_{i_b})$ and $(Y_{i_1}, \ldots, Y_{i_b})$ have statistical distance $\leq \delta$.*

Note that when $\delta = 0$, the variables X_i are exactly dn-wise independent.

Theorem 7 ([HLV17])**.** *Let $f \colon \{0,1\}^m \to \mathbb{C}_{\leq 1}$ be a product test with k functions of input length n. Let D and T be two independent distributions over $\{0,1\}^m$ that are $(0, dn)$-close to uniform. Then*

$$\left|\mathbb{E}[f(D + T \wedge U)] - \mathbb{E}[f(U)]\right| \leq k 2^{-\Omega(d^2 n/k)},$$

where U is the uniform distribution.

Note that the dependence on the number k of functions is poor: when $k = \Omega(d^2 n)$, the error bound does not give anything non-trivial. A main technical contribution of this work is obtaining exponentially better dependency on k using different techniques from [HLV17]. Our theorem gives non-trivial error bound even when $d = O(1)$ and k is exponential in n.

Theorem 8. *Let $f \colon \{0,1\}^m \to \mathbb{C}_{\leq 1}$ be a product test with k functions of input length n. Let D and T be two independent distributions over $\{0,1\}^m$ that are (δ, dn)-close to uniform. Then*

$$\left|\mathop{\mathbb{E}}_{D,T,U}[f(D + T \wedge U)] - \mathop{\mathbb{E}}_{U}[f(U)]\right| \leq 2^{-\Omega(d)} + B\delta,$$

where U is the uniform distribution, for the following choices of B:

i. $B = (k2^n)^{O(d)}$;

ii. *if* f *is nice, then* $B = (d2^n)^{O(d)}$.

Setting $\delta = 0$, Theorem 8 has a better bound than Theorem 7 when $k = \Omega(dn)$. An interesting feature of Theorem 8 is that for nice products the parameter δ can be independent of k. We complement this feature with a negative result showing that for general products a dependence on k is necessary. Thus, the distinction between products and nice products is not an artifact of our proof but is inherent.

Claim 9. *For every sufficiently large* k*, there exists a distribution* D *over* $\{0,1\}^k$ *that is* $(k^{-\Omega(1)}, k^{\Omega(1)})$*-close to uniform, and a product test* $f\colon \{0,1\}^k \to \mathbb{C}_{\le 1}$ *with* k *functions of input length* 1 *such that*

$$\big|\,\mathbb{E}[f(D + T \wedge U)] - \mathbb{E}[f(U)]\big| \ge 1/10,$$

where T *and* U *are the uniform distribution over* $\{0,1\}^k$.

This claim also shows that for $n = 1$ and $\varepsilon = \Omega(1)$ one needs $\delta \le k^{-\Omega(1)}$, and even for distributions which are $(\delta, k^{\Omega(1)})$-close to uniform, instead of just $(\delta, O(1))$-close.

For the class of combinatorial rectangles, which corresponds to product tests with each f_i outputting values in $\{0,1\}$, the classic result [EGL$^+$92] (extended in [CRS00], for an exposition see Lecture 1 in [Vio17]) shows that dn-wise independence alone fools rectangles with error $2^{-\Omega(d)}$ and this error bound is tight. So Theorem 8 does not give better bounds for rectangles, even with the presence of noise. We develop additional machinery and obtain an improvement on Theorem 8. While the improvement is modest, the machinery we develop may be useful for further improvements. Since this improvement is not used in our construction of PRGs, we only state and prove it for exact bounded independence. For technical reasons we restrict the range of the f_i slightly.

Theorem 10. *Let* f *be a product test with* k *functions of input length* n*. Suppose the range of each function* f_i *of* f *is the set* $\{0,1\}$*, or the set of all* M*-th roots of unity for some fixed* M*. Let* D *and* T *be two independent distributions over* $\{0,1\}^m$ *that are* dn*-wise independent. Then*

$$\big|\mathbb{E}[f(D + T \wedge U)] - \mathbb{E}[f(U)]\big| \le n^{-\Omega(d)}.$$

Finally, it is natural to ask if similar techniques fool non-read-once polynomials. In this regard, we are able to show that small-bias distributions [NN93] plus noise fool $\mathbb{F}_2$-polynomials of degree 2.

Claim 11. *Let* $p\colon \{0,1\}^m \to \{0,1\}$ *be any* $\mathbb{F}_2$*-polynomial of degree* 2*. Let* D *and* T *be two distributions over* $\{0,1\}^m$*, where* D *is* δ*-biased, and* T *sets each bit to* 1 *independently with probability* 2/3*. Then*

$$\big|\mathbb{E}[p(D + T \wedge U)] - \mathbb{E}[p(U)]\big| \le \delta.$$

1.1 Techniques

We first explain how to prove bounded independence plus noise fools product tests, i.e., Theorem 8. After that, we will explain the additional ideas that go into constructing our PRGs.

Following the literature, at a high level we do a case analysis based on the *total-variance* of the product test f we want to fool. This is defined as the sum of the variances $\mathrm{Var}[f_i]$ of the functions f_i in the definition of product test. The variance of a function g is $\mathbb{E}[|g(x)|^2] - |\mathbb{E}[g(x)]|^2$ where x is uniform.

Low total-variance. Our starting point is a compelling inequality in [GKM15] (cf. [GMR⁺12, GY14]) showing that bounded independence alone without noise fools low total-variance product tests. However, their result is only proved for *exact* bounded independence, i.e., every d bits are exactly uniform, whereas it is critical for our seed lengths to handle *almost* bounded independence, i.e., every d bits are close to uniform.

One technical contribution in this paper is extending the inequality in [GKM15] to work for almost bounded independence. The proof of the inequality in [GKM15] is somewhat technical, and our extension introduces several complications. For example, the expectations of the f_i under the almost-bounded independent distribution D and the uniform distribution U are not guaranteed to be equal, and this requires additional arguments. However our proof follows the argument in [GKM15], which we also present in a slightly different way that is possibly of interest to some readers. Finally we mention that Claim 9 shows that our error term is close to tight in certain regimes, cf. Section 7.

High total-variance. Here we take a different approach from the ones in the literature: The papers [GLS12, GKM15] essentially reduce the high total-variance case to the low total-variance case. However their techniques either blow up the seed length polynomially [GLS12] or rely on space-bounded generators that only work in fixed order [GKM15].

We instead observe that bounded independence plus noise fools even high total-variance product tests. We now give some details of our approach. A standard fact is that the expectation of a product test f is bounded above by

$$\prod_i |\mathbb{E}[f_i]| \leq \prod_i (1 - \mathrm{Var}[f_i])^{1/2} \leq e^{-\sum_i \mathrm{Var}[f_i]/2}.$$

So if the total-variance $\sum_i \mathrm{Var}[f_i]$ is large then the expectation of the product test under the uniform distribution is small. Thus, it suffices to show that the expectation is also small under bounded independence plus noise. To show this, we argue that typically, the total-variance remains high even considering the f_i as functions of the noise only. Specifically, we first show that on average over a uniform x and t, the variance of the functions $f'_i(y) := f_i(x + t \wedge y)$ is about as large as that of the f_i. This uses Fourier analysis. Then we use concentration inequalities for almost bounded independent distributions to derandomize this fact: we show that it also holds for typical x and t sampled from D and T.

This suffices to prove Theorem 8.i. Proving Theorem 8.ii requires extra ideas.

We first note that the high total-variance case actually does not appear in the read-once CNF generator in [GMR$^+$12]. This is because one can always *truncate* the CNF to have at most $2^w \log(1/\varepsilon)$ number of clauses of width w, which suffices to determine the expected value of the CNF up to an additive error of ε, and such a CNF has low total-variance (for this one argues that noise helps reduce the variance a little.) To handle an arbitrary read-once CNF, [GMR$^+$12] partition the clauses according to their width, and handle each partition separately.

However, one cannot truncate polynomials without noise. To see why, consider, for a simple example, the linear polynomial $x_1 + x_2 + \ldots + x_m$ (corresponding to a product test that computes the parity function). Here no strict subset of the monomials determines the expectation of the polynomial. Indeed, one can construct distributions which look random to $m - 1$ monomials, but not to m.

Truncation using noise. Although we cannot truncate polynomials without noise, we show that something almost as good can still be done, and this idea is critical to obtaining our seed lengths. We show that the statistical closeness parameter in D and T can be selected *as if the polynomial was truncated*: it is independent from the number k of functions. This is reflected in Theorem 8.ii, where δ is independent from k. The proof goes by showing that if the number k of functions is much larger than 2^{2n} then noise alone will be enough to fool the test, regardless of anything else. This proof critically uses noise: without noise a dependence on k is necessary, as shown in the parity example in our discussion. Also, for the proof to work the functions must have expectation at most $1 - \Omega(2^{-n})$. As mentioned earlier, we further prove that this last requirement is necessary (Claim 9): we construct functions whose expectation is about $1-1/k$ but their product is not fooled by almost bounded independence plus noise, if the statistical closeness parameter is larger than $1/k^c$ for a suitable constant c.

Extra ideas for improved bound. To obtain the improved error bound in Theorem 10, we show that whenever the total-variance of a product test lies below $dn^{0.1}$, we can use noise bring it down to below $dn^{-0.1}$. This produces a gap of $[dn^{-0.1}, dn^{0.1}]$ between the high and low total-variance cases, which gives the better bound using the previous arguments. Reducing the total-variance requires a few additional ideas. First, we use Theorem 7 to handle the functions f_i in the product test which have high variances. Then we use the hypercontractivity theorem to reduce the variances of the rest of the f_i individually. [GMR$^+$12] also uses noise to reduce variance, but their f_i are just AND and so they do not need hypercontractivity. To combine both ideas, we prove a new "XOR Lemma" for bounded independence, a variant of an XOR lemma for small-bias, which was proved in [GMR$^+$12].

Constructing our PRGs. We now explain how to use Theorem 8 to construct our PRGs. The high-level idea of our PRG construction is to apply Theorem 8 recursively following the Ajtai–Wigderson framework: Given $D + T \wedge U$, we can think of T as selecting each position in $\{1, \ldots, m\}$ with probability $1/2$. For intuition, it would be helpful to assume each position is selected independently.

We will focus on how to construct a PRG using a seed of length $\tilde{O}(\log m)$ for read-once polynomials with constant error, as this simplifies the parameters and captures all the ideas. Without loss of generality, we can assume the degree of a polynomial to be $n = O(\log m)$, because the contribution of higher-degree terms can be shown to be negligigle under a small-bias distribution (See the proof of Theorem 1.)

Let $p\colon \{0,1\}^m \to \{0,1\}$ be a degree-n read-once polynomial with k monomials. It would be convenient to think of p outputting values $\{-1,1\}$. Further, we can write p as a product $\prod_{i=1}^{k} p_i$, where each p_i is a monomial on at most n bits (with outputs in $\{-1,1\}$.)

Now suppose we only assign the values in D to the positions not chosen by T, that is, setting the input bits $x_i = D_i$ for $i \notin T$. This induces another polynomial $p_{D,T}$ defined on the positions in T. Clearly, $p_{D,T}$ also has degree at most n, and so we can reapply Theorem 8 to $p_{D,T}$.

Repeating the above argument t times induces a polynomial defined on the positions $T_t := \wedge_{i=1}^{t} T_i$. One can think of T_t as a single distribution that selects each position with probability 2^{-t}. Viewing T_t this way, it is easy to see that we can terminate the recursion after $t := O(\log m)$ steps, as the set T_t should become empty with high probability.

By standard constructions [NN93], it takes $s := \tilde{O}(n)$ bits to sample D and T in Theorem 8.ii each time. Therefore, we get a PRG of seed length $t \cdot s = \tilde{O}(n) \log m$.

To obtain a better seed length, we will instead apply Theorem 8 in *stages*. Our goal in each stage is to reduce the degree of the polynomial by half. In other words, we want the restricted polynomial defined on the positions in $\wedge_{i=1}^{t} T_i$ to have degree $n/2$. It is not difficult to see that in order to reduce the degree of the m monomials of p to $n/2$ with high probability, it suffices to apply our above argument recursively for $t := O(\log m)/n$ times. So in each stage, we use a seed of length

$$t \cdot s = \tilde{O}(n) \cdot \Big(\frac{\log m}{n}\Big) = \tilde{O}(\log m).$$

After repeating the same argument for $O(\log n) = \tilde{O}(1)$ stages, with high probability the restricted polynomial would have degree 0 and we are done. Therefore, the total seed length of our PRG is $\tilde{O}(\log m)$.

Here we remark that it is crucial in our argument that D and T are almost-bounded independent, as opposed to being small-biased. Otherwise, we cannot have seed length $s = \tilde{O}(n)$ when $n = o(\log m)$. For example, when $n = O(1)$, with small-bias we would need $s = O(\log m)$ bits, whereas we just use $O(\log \log m)$ bits.

We remark that [GMR$^+$12] uses the *number of clauses* in the CNF as their progress measure. They show that this number drops polynomially at each iteration, allowing them to recurse for only $O(\log \log m)$ steps. However this only holds for truncated CNFs, and is false for example for the linear polynomial $x_1 + x_2 + \ldots + x_m$. This is why we are not able to get a PRG with seed length $\tilde{O}(\log m)$ for polynomially small error.

Organization. We prove bounded independence plus noise fools product (Theorem 8) in Section 2, except the proof of the low total-variance case, which we defer to Section 4. Then

Conditions	Uses	Follows from	Error
(1) $\sum_{i<k} \mathrm{Var}[f_i] \le \alpha d$	D	Lemma 12	$2^{-\Omega(d)} + (k2^n)^{O(d)}\delta$
(2) $\sum_{i<k} \mathrm{Var}[f_i] \ge \alpha d$	$D + T \wedge U$	Derandomized Claim 13	$2^{-\Omega(d)} + k^{O(d)}\delta$
(3) $k \ge 2^{2n+1}d$, nice products	$T \wedge U$	Claim 12	$2^{-\Omega(d)} + 2^{O(dn)}\delta$

Table 1: Error bounds for fooling a product tests of k functions of input length n under different conditions. Here D and T are (δ, dn)-close to uniform, and α is a small constant.

we give constructions of our PRGs in Section 3. In Section 5, we show how to obtain the modest improvement of Theorem 8. After that, we prove our result on fooling degree-2 polynomials in Section 6. Finally, we prove Claim 9 in Section 7.

2 Bounded independence plus noise fools products

In this section we prove Theorem 8. As we mentioned in the introduction, the proof consists of 3 parts: (1) Low total-variance, (2) high total-variance, and (3) truncation using noise for nice products. We summarize the conditions and the error bounds we obtain for these cases in Table 1. Let us now quickly explain how to put them together to prove Theorem 8. Clearly, combining (1) and (2) immediately gives us a bound of $2^{-\Omega(d)} + (k2^n)^{O(d)}$ for product tests, proving Theorem 8.i. For nice product tests, we can apply (3) if $k \ge 2^{2n+1}d$, otherwise we can plug in $k \le 2^{2n+1}d$ in the previous bound, proving Theorem 8.ii.

We now discuss each of the 3 cases in order. Since the proof of the low total-variance case is quite involved, we only state the lemma in this section and defer its proof to Section 4.

Lemma 12. *Let $X_1, X_2, \ldots, X_k$ be k independent random variables over $\mathbb{C}_{\le 1}$ with $\min_{z \in \mathrm{Supp}(X_i)}$ $\Pr[X_i = z] \ge 2^{-n}$ for each $i \in \{1, \ldots, k\}$. Let $Y_1, Y_2, \ldots, Y_k$ be k random variables over $\mathbb{C}_{\le 1}$ that are $(\varepsilon, 16d)$-close to $X_1, \ldots, X_k$. Then*

$$\left| \mathbb{E}\left[\prod_{i=1}^{k} Y_i\right] - \mathbb{E}\left[\prod_{i=1}^{k} X_i\right] \right| \le 2^{O(d)} \left(\frac{\sum_{i=1}^{k} \mathrm{Var}[X_i]}{d} \right)^{d/2} + (k2^n)^{O(d)}\varepsilon.$$

We now prove a claim that handles the high total-variance case. This claim shows that for uniform x and t, the variance of the function $g(y) := f(x + t \wedge y)$ is close to the variance of f in expectation. Its proof follows from a simple calculation in Fourier analysis. Later, we will derandomize this claim in the proof of Theorem 8.

Claim 13. *Let T be the distribution over $\{0,1\}^n$ where the T_j's are independent and $\mathbb{E}[T_j] = \eta$ for each j. Let $f\colon \{0,1\}^n \to \mathbb{C}$ be any function. Then*

$$\mathop{\mathbb{E}}_{U,T}\left[\mathop{\mathrm{Var}}_{U'}[f(U + T \wedge U')]\right] \ge \eta \,\mathrm{Var}[f].$$

Proof of Claim 13. By the definition of variance and linearity of expectation, we have

$$\begin{aligned}\mathop{\mathbb{E}}_{U,T}\Big[\mathop{\mathrm{Var}}_{U'}[f(U+T\wedge U')]\Big] &= \mathop{\mathbb{E}}_{U,T}\Big[\mathop{\mathbb{E}}_{U'}\big[|f(U+T\wedge U')|^2\big] - \Big|\mathop{\mathbb{E}}_{U'}[f(U+T\wedge U')]\Big|^2\Big] \\ &= \mathop{\mathbb{E}}_{U,T}\Big[\mathop{\mathbb{E}}_{U'}\big[|f(U+T\wedge U')|^2\big]\Big] - \mathop{\mathbb{E}}_{U,T}\Big[\Big|\mathop{\mathbb{E}}_{U'}[f(U+T\wedge U')]\Big|^2\Big].\end{aligned}$$

The first term is equal to

$$\mathop{\mathbb{E}}_{U}[|f(U)|^2] = \sum_{\alpha,\alpha'}\hat{f}_\alpha\overline{\hat{f}_{\alpha'}}\mathop{\mathbb{E}}_{U}[\chi_{\alpha-\alpha'}(U)] = \sum_{\alpha}|\hat{f}_\alpha|^2.$$

The second term is equal to

$$\begin{aligned}&\mathop{\mathbb{E}}_{U,T}\Big[\mathop{\mathbb{E}}_{U'}\Big[\sum_{\alpha}\hat{f}_\alpha\chi_\alpha(U+T\wedge U')\Big]\overline{\mathop{\mathbb{E}}_{U''}\Big[\sum_{\alpha'}\hat{f}_{\alpha'}\chi_{\alpha'}(U+T\wedge U'')\Big]}\Big] \\ &= \mathop{\mathbb{E}}_{U,T}\Big[\sum_{\alpha,\alpha'}\hat{f}_\alpha\overline{\hat{f}_{\alpha'}}\mathop{\mathbb{E}}_{U'}[\chi_\alpha(U+T\wedge U')]\mathop{\mathbb{E}}_{U''}[\chi_{\alpha'}(U+T\wedge U'')]\Big] \\ &= \sum_{\alpha,\alpha'}\hat{f}_\alpha\overline{\hat{f}_{\alpha'}}\mathop{\mathbb{E}}_{U}[\chi_{\alpha+\alpha'}(U)]\mathop{\mathbb{E}}_{T}\Big[\mathop{\mathbb{E}}_{U'}[\chi_\alpha(T\wedge U')]\mathop{\mathbb{E}}_{U''}[\chi_{\alpha'}(T\wedge U'')]\Big] \\ &= \sum_{\alpha}|\hat{f}_\alpha|^2\mathop{\mathbb{E}}_{T,U',U''}[\chi_\alpha(T\wedge(U'+U''))] \\ &= \sum_{\alpha}|\hat{f}_\alpha|^2(1-\eta)^{|\alpha|}.\end{aligned}$$

Therefore,

$$\mathop{\mathbb{E}}_{U,T}\Big[\mathop{\mathrm{Var}}_{U'}[f(U+T\wedge U')]\Big] = \sum_{\alpha}|\hat{f}_\alpha|^2\Big(1-(1-\eta)^{|\alpha|}\Big) \geq \eta\sum_{\alpha\neq 0}|\hat{f}_\alpha|^2 = \eta\,\mathrm{Var}[f],$$

where the inequality is because $1-(1-\eta)^{|\alpha|} \geq 1-(1-\eta) \geq \eta$ for any $\alpha \neq 0$. □

With Lemma 12 and Claim 13, we now prove Theorem 8.

Proof of Theorem 8.i. Let σ denote $(\sum_{i\leq k}\mathrm{Var}[f_i])^{1/2}$. We will consider two cases: $\sigma^2 \leq \alpha d$ and $\sigma^2 > \alpha d$, where $\alpha > 0$ is a sufficiently small constant.

If $\sigma^2 \leq \alpha d$, we use Lemma 12. Specifically, since $\Pr[f_i(U) = z] \geq 2^{-n}$ for every $z \in \mathrm{Supp}(f_i)$, it follows from Lemma 12 that

$$\Big|\mathbb{E}\Big[\prod_{i=1}^{k}f_i(D)\Big] - \mathbb{E}\Big[\prod_{i=1}^{k}f_i(U)\Big]\Big| \leq 2^{-\Omega(d)} + (k2^n)^{O(d)}\delta,$$

and the desired bound holds for every fixing of T and U.

If $\sigma^2 \geq \alpha d$, then the expectation of f under the uniform distribution is small. More precisely, we have

$$\Big|\prod_{i\leq k}\mathop{\mathbb{E}}_{U}[f_i(U)]\Big| = \prod_{i\leq k}(1-\mathrm{Var}[f_i])^{1/2} \leq e^{-\frac{1}{2}\sigma^2} \leq 2^{-\Omega(d)}. \tag{1}$$

Thus, it suffices to show that its expectation under $D+T\wedge U$ is at most $2^{-\Omega(d)}+(k2^n)^{O(d)}\delta$. We now use Claim 13 to show that

$$\Big|\mathop{\mathbb{E}}_{D,T,U}\Big[\prod_{i=1}^{k} f_i(D+T\wedge U)\Big]\Big| \leq 2^{-\Omega(d)}+(k2^n)^{O(d)}\delta. \tag{2}$$

For each $t, x \in \{0,1\}^m$, and each $i \in \{1,2,\ldots,k\}$, let $\sigma^2_{t,x,i}$ denote $\mathrm{Var}_{U'}[f_i(x+t\wedge U')]$. We claim that $\sum_{i\leq k}\sigma^2_{t,x,i}$ is large for most x and t sampled from D and T respectively. From Claim 13 we know that this quantity is large in expectation for uniform x and t. By a tail bound for almost bounded independent distributions, we show that the same is true for most $x \in D$ and $t \in T$. By a similar calculation to (1) we show that for these x and t we have that $|\mathbb{E}[f(x+t\wedge U)]|$ is small.

To proceed, let T' be the uniform distribution over $\{0,1\}^m$. Applying Claim 13 with $\eta = 1/2$, we have $\mathbb{E}_{T',U}[\sigma^2_{T',U,i}] \geq \mathrm{Var}[f_i]/2$. So by linearity of expectation,

$$\mathop{\mathbb{E}}_{T',U}\Big[\sum_{i\leq k}\sigma^2_{T',U,i}\Big] \geq \sigma^2/2 \geq \alpha d/2.$$

Since T and D are both (δ, dn)-close to uniform, the random variables $\sigma^2_{T,D,1},\ldots,\sigma^2_{T,D,k}$ are $(2\delta, dn)$-close to $\sigma^2_{T',U,1},\ldots,\sigma^2_{T',U,k}$. Let $\mu = \mathbb{E}_{T',U}[\sum_{i\leq k}\sigma^2_{T',U,i}] \geq \alpha d/2$. By Lemma 48,

$$\Pr_{T',U}\Big[\sum_{i\leq k}\sigma^2_{T',U,i} \leq \mu/2\Big] \leq 2^{-\Omega(d)} + k^{O(d)}\delta.$$

Hence, except with probability $2^{-\Omega(d)}+k^{O(d)}\delta$ over $t \in T$ and $x \in D$, we have

$$\sum_{i\leq k}\sigma^2_{t,x,i} = \sum_{i\leq k}\mathop{\mathrm{Var}}_{U'}[f_i(x+t\wedge U')] \geq \alpha d/4.$$

For every such t and x, we have

$$\begin{aligned}\Big|\prod_{i\leq k}\mathop{\mathbb{E}}_{U}[f_i(x+t\wedge U)]\Big| &\leq \prod_{i\leq k}\Big|\mathop{\mathbb{E}}_{U}[f_i(x+t\wedge U)]\Big| \\ &= \prod_{i\leq k}(1-\sigma^2_{t,x,i})^{1/2} \\ &\leq e^{-\frac{1}{2}\sum_{i\leq k}\sigma^2_{t,x,i}} \leq 2^{-\Omega(d)}. \end{aligned} \tag{3}$$

In addition, we always have $|f| \leq 1$. Hence, summing the R.H.S. of (2) and (3), we have

$$\Big|\mathop{\mathbb{E}}_{D,T,U}\Big[\prod_{i\leq k} f_i(D+T\wedge U)\Big]\Big| \leq \mathop{\mathbb{E}}_{D,T}\Big[\Big|\prod_{i\leq k}\mathop{\mathbb{E}}_{U}[f_i(D+T\wedge U)]\Big|\Big] \leq 2^{-\Omega(d)}+k^{O(d)}\delta. \qquad \square$$

To prove Theorem 8.ii, we use the following additional observation that noise alone fools nice products when k is suitably larger than 2^{2n}. The high-level idea is that in such a case there will be at least $k2^{-n} \geq 2^n$ functions f_i whose inputs are completely set to uniform by the noise. Since the expectation of each f_i is bounded by $1 - O(2^{-n})$, the expectation of their product becomes small when k is suitably larger than 2^{2n}. On the other hand, $\mathbb{E}[f(U)]$ can only get smaller under the uniform distribution, and so the expectations under uniform and noise are both small.

Claim 14 (Noise fools nice products with large k)**.** *Let $f\colon \{0,1\}^m \to \mathbb{C}_{\leq 1}$ be a* nice *product test with $k \geq 2^{2n+1}d$ of input length n. Let T be a distribution over $\{0,1\}^m$ that is (δ, dn)-close to uniform. Then*

$$\left|\mathop{\mathbb{E}}_{T,U}[f(T \wedge U)] - \mathbb{E}[f(U)]\right| \leq 2^{-\Omega(d)} + 2^{O(dn)}\delta.$$

Proof. We will bound above both expectations in absolute value. Let $k' := 2^{2n+1}d \leq k$. Write $f = \prod_{i=1}^{k} f_i$, where $f_i\colon \{0,1\}^{I_i} \to \mathbb{C}_{\leq 1}$. Since f is nice, there is a constant $\alpha \in (0,1]$ such that $|\mathbb{E}[f_i(U)]| \leq 1 - \alpha 2^{-n}$ for every $i \in \{1, \ldots, k\}$. Under the uniform distribution, we have

$$\left|\mathbb{E}[f(U)]\right| = \prod_{i=1}^{k}\left|\mathbb{E}[f_i(U)]\right| \leq (1 - \alpha 2^{-n})^k \leq e^{-\Omega(k2^{-n})} \leq 2^{-\Omega(d)}. \tag{4}$$

It suffices to show that the expectation under $T \wedge U$ is at most $2^{-\Omega(d)} + 2^{O(dn)}\delta$. Note that

$$\left|\mathbb{E}[f(T \wedge U)]\right| \leq \mathop{\mathbb{E}}_{T}\Big[\prod_{i=1}^{k}\big|\mathop{\mathbb{E}}_{U}[f_i(T \wedge U)]\big|\Big] \leq \mathop{\mathbb{E}}_{T}\Big[\prod_{i=1}^{k'}\big|\mathop{\mathbb{E}}_{U}[f_i(T \wedge U)]\big|\Big].$$

We now show that the R.H.S. is at most $2^{-\Omega(d)} + 2^{O(dn)}\delta$. We first show that the expected number of f_i whose inputs are all selected by T when T is uniform is large, and then apply a tail bound for almost bounded independent distributions to show that it holds for most $t \in T$. Let T' be the uniform distribution over $\{0,1\}^m$. Then

$$\mathbb{E}\Big[\sum_{i=1}^{k'} \mathbb{1}(T'_{I_i} = 1^{|I_i|})\Big] = \sum_{i=1}^{k'} \Pr[T'_{I_i} = 1^{|I_i|}] \geq k'2^{-n} = 2^{n+1}d.$$

Since T is (δ, dn)-close to uniform, the T_{I_i} are (δ, d)-close to uniform. By Lemma 48,

$$\Pr_{T}\Big[\sum_{i=1}^{k'} \mathbb{1}(T_{I_i} = 1^{|I_i|}) \leq 2^n d\Big] \leq 2^{-\Omega(dn)} + 2^{O(dn)}\delta. \tag{5}$$

Note that if $T_{I_i} = 1^{|I_i|}$, then $|\mathbb{E}_U[f_i(T \wedge U)]| = |\mathbb{E}[f]| \leq 1 - \alpha 2^{-n}$. Thus, conditioned on $\sum_{i=1}^{k'} \mathbb{1}(T_{I_i} = 1^{|I_i|}) \geq 2^n d$, we have

$$\prod_{i=1}^{k'}\left|\mathbb{E}[f_i(T \wedge U)]\right| \leq (1 - \alpha 2^{-n})^{2^n d} \leq 2^{-\Omega(d)}. \tag{6}$$

Since we always have $|f| \leq 1$, the error bound follows from summing the R.H.S. of (4), (5) and (6). □

Theorem 8.ii now follows easily from Claim 14 and Theorem 8.i.

Proof of Theorem 8.ii. Since f is nice, there is a constant $\alpha \in (0, 1]$ such that $|\mathbb{E}[f_i]| \leq 1 - \alpha 2^{-n}$. If $k \geq 2^{2n+1}d$, then the theorem follows from Claim 14. Otherwise, $k \leq 2^{2n+1}d$ and the theorem follows from Theorem 8.i. □

3 Pseudorandom generators

In this section we construct our generators. As explained in the introduction, all constructions follow from applying the Theorem 8 recursively. The following lemma captures the trade-off between the number of recursions and the simplification on a product test. As a first read, we suggest the readers to refer to the $\tilde{O}$ notations in the statements and proofs, i.e., ignore polylogarithmic factors in n, $\log k$, $\log(1/\varepsilon)$ and $\log m$, and think of k as m and ε as some arbitrary small constant.

Lemma 15. *If there is an explicit generator $G' \colon \{0,1\}^{\ell'} \to \{0,1\}^m$ that fools product tests with k functions of input length r with error ε' and seed length ℓ', then there is an explicit generator $G \colon \{0,1\}^{\ell} \to \{0,1\}^m$ that fools product tests with k functions of input length n with error $\varepsilon' + t\varepsilon$, where $t = O\big(\frac{\log(k/\varepsilon)}{r+1} + \log(\frac{n}{r+1})\big) = \tilde{O}\big(\frac{\log(k/\varepsilon)}{r+1} + 1\big)$, and seed length*

i. $\ell = \ell' + t \cdot O((\log k + n)\log(1/\varepsilon) + \log\log m) = \ell' + t \cdot \tilde{O}((\log k + n)\log(1/\varepsilon))$;

ii. if the product tests are nice, then $\ell = \ell' + t \cdot O((n + \log\log(1/\varepsilon))\log(1/\varepsilon) + \log\log m) = t \cdot \tilde{O}(n\log(1/\varepsilon))$.

We defer its proof to the end. Our generator for arbitrary product tests (Theorem 3) easily follows from the lemma. Our generator for nice product tests (Theorem 5) requires applying the lemma in stages, where in each stage we apply the lemma with a different value of n. XORing its output with a small-bias distribution gives our generator for polynomials (Theorem 1).

Proof of Theorem 3. We apply Lemma 15.i with $r = 0$ with error ε/t, where $t = O(\log(k/\varepsilon) + \log n) = \tilde{O}(\log(k/\varepsilon))$. Note that a product test of input length 0 is a constant function, which can always be fooled with zero error. So we have a generator that fools product tests with k functions of input length n, with error ε and seed length

$$t \cdot O\left(n\log(t/\varepsilon) + \log(1/\delta) + \log\log m\right) = \tilde{O}(\log(k/\varepsilon))(n + \log k)\log(1/\varepsilon).$$ □

We will apply Lemma 15 in $O(\log n)$ stages. In each stage our goal is to halve the input length of the product test.

Proof of Theorem 5. Let f be a nice product test with k functions of input length n. Note that by applying Lemma 15 with $r = n/2$ and error $\varepsilon/(t \log n)$, where $t = O(\log(k/\varepsilon)/n + 1)$, we can halve its input length by incurring an error of $\varepsilon/O(\log n)$ and using a seed of length

$$\begin{aligned} t \cdot O\big(n \log(t \log n/\varepsilon) + \log(1/\delta) + \log\log m\big) &= t \cdot \tilde{O}\big(n \log(1/\varepsilon) + \log(1/\delta)\big) \\ &= \tilde{O}(\log(k/\varepsilon) + n) \log(1/\varepsilon). \end{aligned}$$

Now we repeat the argument for $s = O(\log n)$ steps until the input length is zero, which is a constant function and can be fooled with zero error. So we have a generator that fools nice product tests with k functions of input length n, with error ε and seed length $s \cdot \tilde{O}(\log(k/\varepsilon) + n) \log(1/\varepsilon) = \tilde{O}(\log(k/\varepsilon) + n) \log(1/\varepsilon)$. □

Theorem 1 follows from XORing the output of the above generator with a small-bias distribution.

Proof of Theorem 1. Let c be a sufficiently large constant. Let D be a $(\varepsilon/m)^c$-biased distribution over $\{0,1\}^m$ [NN93]. Let G be the output distribution of the generator in Theorem 5 that fools product tests with m functions and input length $c \log(m/\varepsilon)$ with $\varepsilon/2$. The generator outputs $D + G$. By [NN93] and Theorem 5, it takes a seed of length

$$O(\log(m/\varepsilon)) + \tilde{O}\big(\log(m/\varepsilon) + c \log(m/\varepsilon)\big) \log(1/\varepsilon) = \tilde{O}(\log(m/\varepsilon)) \log(1/\varepsilon).$$

Let $p\colon \{0,1\}^m \to \{-1,1\}$ be any read-once GF(2) polynomial. Consider the polynomial p' obtained from p by removing all the monomials with degree greater than $c \log(m/\varepsilon)$ in p. We claim that the expectation of p and p' under D differs by at most ε. Note that under any $(\varepsilon/m)^c$-biased distribution X, the probability that any $c \log(m/\varepsilon)$ bits are 1 is at most $\varepsilon/4m$, and so by a union bound we have $\Pr[p(X) \neq p'(X)] \leq \varepsilon/4$. In particular, this holds for D and U. It follows that

$$\big|\mathbb{E}[p(D+G)] - \mathbb{E}[p(U)]\big| \leq \big|\mathbb{E}[p'(D+G)] - \mathbb{E}[p'(U)]\big| + \varepsilon/2 \leq \varepsilon,$$

where the last inequality holds for any fixed D because of Theorem 5. □

We now prove Lemma 15. First we state a claim that will be used in the proof to reduce the input length of the product test.

Claim 16. *Let $T^{(1)}, \dots, T^{(t)}$ be t independent and identical distributions over $\{0,1\}^n$ that are δ-close to uniform. Then $\Pr[wt(\wedge_{i=1}^t T^{(i)}) > r] \leq \binom{n}{r+1}(2^{-(r+1)} + \delta)^t$.*

Proof. Since $T^{(1)}, \dots, T^{(t)}$ are independent and each $T^{(i)}$ is δ-close to uniform,

$$\begin{aligned} \Pr\big[wt(\wedge_{i=1}^t T^{(i)}) > r\big] &\leq \sum_{S:|S|=r+1} \Pr\Big[\wedge_{i=1}^t \wedge_{j\in S} (T_j^{(i)} = 1)\Big] \\ &= \sum_{S:|S|=r+1} \prod_{i=1}^t \Pr\big[\wedge_{j\in S}(T_j^{(i)} = 1)\big] \\ &\leq \sum_{S:|S|=r+1} (2^{-(r+1)} + \delta)^t = \binom{n}{r+1}(2^{-(r+1)} + \delta)^t. \end{aligned}$$

□

Proof of Lemma 15. For $S \subseteq \{1,2,\ldots,m\}$, define the function $\mathrm{PAD}_S(x) \colon \{0,1\}^{|S|} \to \{0,1\}^m$ to output m bits of which the positions in S are the first $|S|$ bits of $x0^{|S|}$ and the rest are 0.

Let C be a sufficiently large constant. The generator G will output $H^{(1)}$, where we define the distribution $H^{(i)}$ recursively for $t = O\big(\frac{\log(k/\varepsilon)}{r+1} + \log(\frac{n}{r+1})\big)$ steps: At the i-th step, $H^{(i)}$ samples two independent distributions $D^{(i)}, T^{(i)}$ over $\{0,1\}^m$ that are $(\delta, Cn\log(1/\varepsilon))$-close to uniform, where δ is taken to be

1. $\delta = 2^{-C(\log k + n)\log(1/\varepsilon)}$;
2. if the product tests are nice, then $\delta = 2^{-C(n + \log\log(1/\varepsilon))\log(1/\varepsilon)}$.

Then output

$$H^{(i)} := D^{(i)} + T^{(i)} \wedge \mathrm{PAD}_{T^{(i)}}(H^{(i+1)}).$$

We define $H^{(t+1)}$ to be $G'(U_{\ell'})$.

By [NN93, Lemma 4.2], sampling $D^{(i)}$ and $T^{(i)}$ takes a seed of length $O(n\log(1/\varepsilon) + \log(1/\delta) + \log\log m)$. The total seed length of G is therefore $\ell = \ell' + t \cdot O(n\log(1/\varepsilon) + \log(1/\delta) + \log\log m)$.

We now analyze the error of G. For $i \in \{1,2,\ldots,t\}$, consider the variant $H_U^{(i)}$ of $H^{(1)}$, which is the same as $H^{(1)}$ but at the i-th step replace $\mathrm{PAD}_{T^{(i)}}(H^{(i+1)})$ with $\mathrm{PAD}_{T^{(i)}}(U_m)$. Let $H_U^{(0)} = U_m$.

For every $i \in \{1,\ldots,t\}$, for every fixed $D^{(1)},\ldots,D^{(i-1)}$ and $T^{(1)},\ldots,T^{(i-1)}$, the function f restricted to $\wedge_{j<i} T^{(j)}$ remains a product test with k functions of input length n, and remains nice if f is nice. Call the restricted function g. Then, by Theorem 8, we have

$$\big|\mathbb{E}[f(H_U^{(i-1)})] - \mathbb{E}[f(H_U^{(i)})]\big| = \big|\mathbb{E}[g(U)] - \mathbb{E}[g(D^{(i)} + T^{(i)} \wedge U_m)]\big| \le \varepsilon.$$

Hence, summing over i we have

$$\big|\mathbb{E}[f(U_m)] - \mathbb{E}[f(H_U^{(t)}]\big| \le \sum_{i=1}^{t} \big|\mathbb{E}[f(H_U^{(i-1)})] - \mathbb{E}[f(H_U^{(i)})]\big| \le t\varepsilon.$$

We now prove that $|\mathbb{E}[f(H_U^{(t)})] - \mathbb{E}[f(H^{(1)})]| \le \varepsilon' + 2\varepsilon$. We will show that except with probability ε, the function f restricted to $\wedge_{j\le t} T^{(j)}$ is a product test of input length r and so we can fool the restricted function using G' given by our assumption.

Write $f = \prod_{i\le k} f_i$, where each f_i is defined on $\{0,1\}^{I_i}$ with $|I_i| \le n$. We claim that

$$\Pr\Big[wt(\wedge_{i=1}^{t} T_{I_j}^{(i)}) > r \text{ for some } j \in \{1,\ldots,k\}\Big] \le \varepsilon.$$

It suffices to analyze $\Pr[wt(\wedge_{i=1}^{t} T_{I_j}^{(i)}) > r]$ for each j and take a union bound over $j \le k$.

Since $|I_j| \le n$, $T_{I_j}^{(i)}$ is 2^{-Cn}-close to uniform, by Claim 16 and a union bound over $j \le k$, the probability that some f_i has input length $> r$ is at most

$$k\binom{n}{r+1}\big(2^{-(r+1)} + 2^{-Cn}\big)^t \le k \cdot \left(\frac{ne}{r+1}\right)^{r+1} \big(2^{-r}\big)^{\Omega\left(\frac{\log(k/\varepsilon)}{r+1} + \log(\frac{n}{r+1})\right)} \le \varepsilon.$$

Hence, for every $D^{(1)}, \dots, D^{(t)}$, with probability $1 - \varepsilon$ over the choice of $T^{(1)}, \dots, T^{(t)}$, the function f restricted to $\wedge_{i=1}^{t} T^{(i)}$ becomes a product with k functions of input length r, and remains nice if f is nice. Conditioned on this, we have by the definition of G' that $|\mathbb{E}[f(H_U^{(t)})] - \mathbb{E}[f(H^{(1)})]| \leq \varepsilon'$. Otherwise, as $|f|$ is bounded by 1, the absolute difference is always at most 2. Hence, $|\mathbb{E}[f(H_U^{(t)})] - \mathbb{E}[f(H^{(1)})]| \leq \varepsilon' + 2\varepsilon$, and so the total error is at most $\varepsilon' + (t + 2)\varepsilon$. □

4 On almost k-wise independent variables with small total-variance

In this section we will prove Lemma 12. Our proof follows closely to the one in [GKM15], which proves the lemma for $\varepsilon = 0$, that is, when the X_i's are d-wise independent. We first give an overview of their proof.

For independent random variables $Z_1, \dots, Z_k$, we will use $\sigma(Z)$ to denote the standard deviation of $\sum_{i \leq k} Z_i$, that is, $\sigma(Z) := (\sum_{i=1}^{k} \mathrm{Var}[Z_i])^{1/2}$.

As a first step, let us assume each $\mathbb{E}[X_i]$ is nonzero and normalize the variables X_i by writing

$$\prod_i X_i = \prod_i (\mathbb{E}[X_i] + (X_i - \mathbb{E}[X_i])) = \prod_i \mathbb{E}[X_i] \cdot \prod_i \left(1 + \frac{X_i - \mathbb{E}[X_i]}{\mathbb{E}[X_i]}\right).$$

Let Z_i denote $(X_i - \mathbb{E}[X_i]) / \mathbb{E}[X_i]$. If $|Z_i|$ is small, then intuitively a low-order Taylor's expansion of $\prod_i (1 + Z_i)$ should approximate the original function well. To write down its Taylor's expansion, a convenient way is to rewrite $\prod_i (1 + Z_i)$ as $e^{\sum_i \log(1+Z_i)}$. It suffices to bound above its error term in expectation. This is equivalent to bounding the d-th moment of $\sum_i \log(1 + Z_i)$. A standard calculation gives a bound in terms of the norm and variance of the functions $\log(1 + Z_i)$. Since $|Z_i|$ is small, $\log(1 + Z_i)$ behaves similarly as Z_i. So we can relate the error term in terms of $|Z_i|$ and $\sigma(Z)^2 := \sum_i \mathrm{Var}[Z_i]$. In particular if $|Z_i| \leq B$ for all i then we would get an error bound of the form $2^{O(d)}(\sqrt{\sigma(Z)^2/d} + B)^{O(d)}$. For now let's think of $\mathbb{E}[X_i]$ being bounded away from 0 so that $\mathrm{Var}[Z_i] = \Theta(\mathrm{Var}[X_i])$.

Now we handle the case where $|Z_i|$ is large. Note that this implies either (1) $|X_i - \mathbb{E}[X_i]|$ is large, or (2) $\mathbb{E}[X_i]$ is small. We will handle the two conditions separately by a reduction to the case where the $|Z_i|$'s are small.

The recurring idea throughout is that we can always tolerate $O(d)$ bad variables that violates the conditions, provided with high probability there can be at most $O(d)$ bad variables. This is because by affording an extra $O(d)$ amount of independence in the beginning, we can condition on the values of these variables and work with the remaining ones.

As a simple illustration of this idea, throughout the proof we can assume each $\mathrm{Var}[X_i]$ is bounded by $\sum_i \mathrm{Var}[X_i]/d =: \sigma(X)^2/d$, as there can be at most d bad variables X_i that violate this inequality, and so we can start with $2d$-wise independence, then conditioned on values of the bad variables X_i, the rest of the X_i would satisfy the bound.

We first assume the $|\mathbb{E}[X_i]|$'s are large and handle (1), we will round the X_i to $\mathbb{E}[X_i]$ whenever $|X_i - \mathbb{E}[X_i]| \geq B$. Note that by Chebyshev's inequality an X_i gets rounded with probability $\mathrm{Var}[X_i]/B^2$. It follows that the probability that there are more than d such X_i's is bounded by $(\sigma(X)/Bd)^d$. This suggests taking B to be $(\sigma(X)/d)^{\alpha}$ for some constant $\alpha \in (0, 1)$ to balance the error terms.

It remains to handle condition (2), for Z_i to be bounded by $B = (\sigma(X)^2/d)^{\Omega(1)}$, as explained above it suffices to show that all but $O(d)$ of the X_i's satisfy $|\mathbb{E}[X_i]| \geq (\sigma(X)^2/d)^{O(1)}$. If $|\mathbb{E}[X_i]| \geq (\sigma(X)/d)^{\Omega(1)}$ for $\Omega(d)$ of the X_i's, then by a similar argument as above one can show that with high probability at least half of them is bounded by $(\sigma(X)^2/d)^{\Omega(1)}$. Hence, $\mathbb{E}[\prod_i X_i]$ is at most $(\sigma(X)^2/d)^{\Omega(d)}$ when the X_i's are d-wise independent. This finishes the proof.

Note that in the case of $\varepsilon > 0$, each X_i is only ε-close to the corresponding Y_i and they are not exactly identical. As a result, throughout the proof we will often have to introduce hybrid terms to move from functions of X_i to functions of Y_i, and vice versa, and we will show that each of these steps introduces an error of at most $k^{O(d)}\varepsilon$.

Also, there is some loss in ε whenever we condition on the values of any subset of the Y_i's, see Claim 24 for a formal claim. This introduces the extra condition that each X_i must put a certain mass on each outcome.

4.1 Preliminaries

In this section, we prove several claims that will be used in the proof of Lemma 12.

Lemma 17. *For any $z \in \mathbb{C}$ with $|z| \leq 1/2$, $|\log(1+z)| \leq 2|z|$, where we take the principle branch of the logarithm.*

Proof. From the Taylor series expansion of the complex-valued log function we have

$$|\log(1+z)| = \left| \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n!} z^n \right| \leq \sum_{n=1}^{\infty} |z|^n \leq |z| \sum_{n=0}^{\infty} (1/2)^n = 2|z|. \qquad \square$$

Lemma 18. *Let $Z \in \mathbb{C}$ be a random variable with $|Z| \leq 1/2$, $\mathbb{E}[Z] = 0$ and $W = \log(1+Z)$ the principle branch of the logarithm function (phase between $(-\pi, \pi)$). We have $\mathrm{Var}[W] \leq 4\,\mathrm{Var}[Z]$.*

Proof. By the definition of Variance, Lemma 17, and that $\mathbb{E}[Z] = 0$,

$$\begin{aligned} \mathrm{Var}[W] &= \mathbb{E}[|W|^2] - |\mathbb{E}[W]|^2 \\ &\leq \mathbb{E}[|W|^2] \\ &\leq 4\,\mathbb{E}[|Z|^2] \\ &= 4\,\mathrm{Var}[Z]. \end{aligned} \qquad \square$$

Lemma 19 (Taylor's approximation)**.** *For $w \in \mathbb{C}$ and $d > 0$,*

$$\left| e^w - \sum_{j=0}^{d-1} w^j/j! \right| \leq O(1) \frac{|w|^d}{d!} \cdot \max\{1, e^{\Re(w)}\}.$$

Lemma 20. *For any random variable $W \in \mathbb{C}$, $|e^{\mathbb{E}[W]}| \le \mathbb{E}[|e^W|]$.*

Proof. By Jensen's inequality, we have

$$|e^{\mathbb{E}[W]}| = |e^{\mathbb{E}[\Re(W)]}| \le |\mathbb{E}[e^{\Re(W)}]| = \mathbb{E}[|e^W|]. \qquad \square$$

Claim 21. $|e^{z_1} - e^{z_2}| \le |e^{z_2}| \cdot O(|z_1 - z_2|)$ *if* $|z_1 - z_2| \le 1$*,*

Proof. By Lemma 19 with $d = 1$,

$$|e^{z_1 - z_2} - 1| \le O(1) \cdot |z_1 - z_2| \cdot \max\{1, e^{\Re(z_1 - z_2)}\} = O(|z_1 - z_2|),$$

because $\Re(z_1 - z_2) \le |z_1 - z_2| \le 1$. Therefore,

$$\begin{aligned} |e^{z_1} - e^{z_2}| &= |e^{z_2}(e^{z_1 - z_2} - 1)| \\ &= |e^{z_2}||e^{z_1 - z_2} - 1| \\ &\le |e^{z_2}| \cdot O(|z_1 - z_2|). \end{aligned} \qquad \square$$

Claim 22. *Let $X, Y \in \Omega$ be two discrete random variables such that $\mathrm{sd}(X, Y) \le \varepsilon$. Let $f \colon \Omega \to \mathbb{C}$ be any function. We have $|\mathbb{E}[f(X)] - \mathbb{E}[f(Y)]| \le 2 \max_z |f(z)| \cdot \mathrm{sd}(X, Y)$.*

Proof. Let p and q be the probability function of X and Y. Using the fact that $\mathrm{sd}(X, Y) = \frac{1}{2} \sum_z |p(z) - q(z)|$, we have

$$\begin{aligned} \Big|\mathbb{E}[f(X)] - \mathbb{E}[f(Y)]\Big| &= \Big|\sum_z p(z)f(z) - \sum_z q(z)f(z)\Big| \\ &\le \sum_z |f(z)||p(z) - q(z)| \\ &\le \max_z |f(z)| \cdot \sum_z |p(z) - q(z)| \\ &= 2 \max_z |f(z)| \cdot \mathrm{sd}(X, Y). \end{aligned} \qquad \square$$

Claim 23 (Maclaurin's inequality (cf. [Ste04]))**.** *Let $z_1, \ldots, z_k$ be k non-negative numbers. For any $i \in \{0, \ldots, k\}$, we have*

$$S_i(z_1, \ldots, z_k) := \sum_{S : |S| = i} \prod_{j \in S} z_j \le (e/i)^i (\sum_{j=1}^k z_j)^i.$$

4.2 Proof of Lemma 12

We now prove Lemma 12. For independent random variables $Z_1, \ldots, Z_k$, we will use $\sigma(Z)$ to denote the standard deviation of $\sum_{i \le k} Z_i$, that is, $\sigma(Z) := (\sum_{i=1}^k \mathrm{Var}[Z_i])^{1/2}$. We will also denote $\sigma(Z)^2/d$ by v for notational simplicity.

4.2.1 Assuming the variances are not too small

As hinted in the overview above, throughout the proof we will without loss of generality assume $\mathrm{Var}[X_i] \leq \sigma(X)^2/d$ for every $i \in \{1, \ldots, k\}$. This assumption will be used in the proof of Lemma 30 to give a uniform bound on how close the rounded X_i's and X_i's are in expectation.

We first prove a claim that shows the Y_i's remains close to the X_i even if we condition on the values of a few of the Y_i's. This claim will be used multiple times throughout the proof. Note that this claim is immediate for exact independence ($\varepsilon = 0$) but less for almost independence. We shall use the assumption that the X_i take any value with probability at least 2^{-n}.

Claim 24. *Let $X_1, X_2, \ldots, X_k$ be k independent random variables over $\mathbb{C}_{\leq 1}$ with $\min_{z \in \mathrm{Supp}(X_i)} \Pr[X_i = z] \geq 2^{-n}$. Let $Y_1, Y_2, \ldots, Y_k$ be k random variables over $\mathbb{C}_{\leq 1}$ that are (ε, d)-close to $X_1, X_2, \ldots, X_k$. Let $S \subseteq \{1, \ldots, k\}$ be a subset of size t. Then conditioned on any values of the Y_i for $i \in S$, the Y_i for $i \notin S$ are $(3 \cdot 2^{2tn}\varepsilon, d - t)$-close to the X_i for $i \notin S$.*

Proof. Let $T \subseteq [k] - S$ be a subset of size at most $d - t$. We have for any value z_ℓ for $\ell \in S$,

$$\sum_{z_j : j \in T} \left| \Pr\Big[\bigwedge_{j \in T} Y_j = z_j \mid \bigwedge_{\ell \in S} Y_\ell = z_\ell\Big] - \Pr\Big[\bigwedge_{j \in T} X_j = z_j\Big] \right|$$
$$= \sum_{z_j : j \in T} \left| \frac{\Pr\Big[\bigwedge_{j \in S \cup T} Y_j = z_j\Big]}{p_Y} - \frac{\Pr\Big[\bigwedge_{j \in S \cup T} X_j = z_j\Big]}{p_X} \right|,$$

where $p_X := \Pr[\wedge_{\ell \in S} X_\ell = z_\ell]$ and $p_Y := \Pr[\wedge_{\ell \in S} Y_\ell = z_\ell]$. Hence, we can rewrite above as

$$\sum_{z_j : j \in T} \left| \Big(\frac{1}{p_Y} - \frac{1}{p_X}\Big) \Pr\Big[\bigwedge_{j \in S \cup T} Y_j = z_j\Big] + \frac{1}{p_X}\Big(\Pr\Big[\bigwedge_{j \in S \cup T} Y_j = z_j\Big] - \Pr\Big[\bigwedge_{j \in S \cup T} X_j = z_j\Big]\Big) \right|$$
$$\leq \left| \frac{1}{p_Y} - \frac{1}{p_X} \right| \sum_{z_j : j \in T} \Pr\Big[\bigwedge_{j \in S \cup T} Y_j = z_j\Big] + \frac{\varepsilon}{p_X}$$
$$\leq |1/p_Y - 1/p_X| + \varepsilon/p_X$$
$$\leq (1/p_X p_Y + 1/p_X)\varepsilon.$$

The first and last inequalities are because the X_i's are (ε, d)-close to the Y_i's. As the X_i's are independent, by our assumption we have $p_X = \prod_{\ell \in S} \Pr[X_\ell = z_\ell] \geq 2^{-tn}$, and so $p_Y \geq 2^{-tn} - \varepsilon \geq 2^{-tn}/2$. (Otherwise the conclusion is trivial.) Therefore, $(1/p_X p_Y + 1/p_X)\varepsilon \leq 3 \cdot 2^{2tn}\varepsilon$, and the proof follows. □

Claim 25. *Let $X_1, X_2, \ldots, X_k$ be k independent random variables over $\mathbb{C}_{\leq 1}$ with $\min_{z \in \mathrm{Supp}(X_i)} \Pr[X_i = z] \geq 2^{-n}$ for each $i \in \{1, \ldots, k\}$. Let $Y_1, Y_2, \ldots, Y_k$ be k random variables over $\mathbb{C}_{\leq 1}$. If Lemma 12 holds when the Y_i's are (Cd, ε)-close to the X_i's assuming $\mathrm{Var}[X_i] \leq \sigma(X)^2/d$ for every $i \in [k]$, then Lemma 12 holds when the Y_i's are $((C+1)d, \varepsilon)$-close the X_i's without the assumption.*

Proof. Note that there can be at most d different such indices. Let J be the set of these indices. We have

$$\begin{aligned}\prod_i X_i - \prod_i Y_i &= \prod_{j\in J} X_j \prod_{i\notin J} X_i - \prod_{j\in J} Y_j \prod_{i\notin J} Y_i \\ &= \Big(\prod_{j\in J} X_j - \prod_{j\in J} Y_j\Big) \prod_{i\notin J} X_j + \prod_{j\in J} Y_j \Big(\prod_{i\notin J} X_j - \prod_{i\notin J} Y_j\Big).\end{aligned}$$

We first bound the expectation of the first term. Since the X_i's are independent,

$$\begin{aligned}\left|\mathop{\mathbb{E}}_{X,Y}\Big[\Big(\prod_{j\in J} X_j - \prod_{j\in J} Y_j\Big)\prod_{i\notin J} X_j\Big]\right| &= \left|\mathbb{E}\Big[\prod_{j\in J} X_j\Big] - \mathbb{E}\Big[\prod_{j\in J} Y_j\Big]\right| \cdot \left|\mathbb{E}\Big[\prod_{i\notin J} X_j\Big]\right| \\ &\le \left|\mathbb{E}\Big[\prod_{j\in J} X_j\Big] - \mathbb{E}\Big[\prod_{j\in J} Y_j\Big]\right| \\ &\le \varepsilon.\end{aligned}$$

For the second term, note that conditioning on the values of the Y_j for which $j \in J$, by Claim 24, the remaining variables are $(2^{O(dn)}\varepsilon, Cd)$-close to the corresponding X_j's. So we can apply the above Lemma 12 with our assumption and the claim follows. □

4.2.2 Assuming the variables are close to their expectations and the expectations are large

Lemma 26. *Let $X_1, X_2, \ldots, X_k$ be k independent discrete random variables over $\mathbb{C}_{\le 1}$. Let $Y_1, Y_2, \ldots, Y_k$ be k discrete random variables over $\mathbb{C}_{\le 1}$ that are (ε, d)-close to $X_1, \ldots, X_k$. Assume for each X_i and Y_i, there exist Z_i and Z'_i such that*

$$X_i = \mathbb{E}[X_i](1 + Z_i) \quad \textit{and} \quad Y_i = \mathbb{E}[X_i](1 + Z'_i),$$

where $|Z_i| \le B \le 1/2$ and $|Z'_i| \le B \le 1/2$. Then

$$\left|\mathbb{E}\Big[\prod_{i=1}^k X_i\Big] - \mathbb{E}\Big[\prod_{i=1}^k Y_i\Big]\right| \le 2^{O(d)} \left(\frac{\sigma(Z)\sqrt{d} + Bd}{d}\right)^d + (Bk)^{O(d)}\varepsilon.$$

Remark 27. *Note that we define Y_i above in terms of $\mathbb{E}[X_i]$ but not $\mathbb{E}[Y_i]$. The Z_i's are independent, but the Z_i's may not be. Also, later we will take B to be $v^{1/3}$.*

Proof. Define $W_i, \hat{W}_i$ such that

$$W_i = \log(1 + Z_i) \quad \text{and} \quad \hat{W}_i = W_i - \mathbb{E}[W_i].$$

Likewise, define $W'_i, \hat{W}'_i$ such that

$$W'_i = \log(1 + Z'_i) \quad \text{and} \quad \hat{W}'_i = W'_i - \mathbb{E}[W'_i].$$

Let $\hat{W} = \sum_i \hat{W}_i$ and $\hat{W}' = \sum_i \hat{W}_i'$. Note that $X_i = \mathbb{E}[X_i] e^{\hat{W}_i + \mathbb{E}[W_i]}$ and $Y_i = \mathbb{E}[Y_i] e^{\hat{W}_i' + \mathbb{E}[W_i']}$. We have

$$\prod_{i=1}^k X_i = \Big(\prod_{i=1}^k \mathbb{E}[X_i] e^{\mathbb{E}[W_i]}\Big) e^{\hat{W}} \quad \text{and} \quad \prod_{i=1}^k Y_i = \Big(\prod_{i=1}^k \mathbb{E}[X_i] e^{\mathbb{E}[W_i']}\Big) e^{\hat{W}'}.$$

Hence the difference is

$$\begin{aligned} \prod_{i=1}^k X_i - \prod_{i=1}^k Y_i &= \Big(\prod_{i=1}^k \mathbb{E}[X_i]\Big)\Big(\prod_{i=1}^k e^{\mathbb{E}[W_i]} \cdot e^{\hat{W}} - \prod_{i=1}^k e^{\mathbb{E}[W_i']} \cdot e^{\hat{W}'}\Big) \\ &= \Big(\prod_{i=1}^k \mathbb{E}[X_i]\Big) \left(\Big(\prod_{i=1}^k e^{\mathbb{E}[W_i]} - \prod_{i=1}^k e^{\mathbb{E}[W_i']}\Big) e^{\hat{W}} + \prod_{i=1}^k e^{\mathbb{E}[W_i']} \cdot \Big(e^{\hat{W}} - e^{\hat{W}'}\Big)\right). \end{aligned}$$

The lemma follows from the two claims below:

Claim 28. *For every outcome of* $\hat{W}$, $\left|\Big(\prod_{i=1}^k \mathbb{E}[X_i]\Big)\Big(\prod_{i=1}^k e^{\mathbb{E}[W_i]} - \prod_{i=1}^k e^{\mathbb{E}[W_i']}\Big) e^{\hat{W}}\right| \le O(k\varepsilon)$.

Claim 29. $\left|\Big(\prod_{i=1}^k \mathbb{E}[X_i] e^{\mathbb{E}[W_i']}\Big)\Big(\mathbb{E}[e^{\hat{W}}] - \mathbb{E}[e^{\hat{W}'}]\Big)\right| \le 2^{O(d)} \left(\frac{\sigma(Z)\sqrt{d} + Bd}{d}\right)^d + (Bk)^{O(d)}\varepsilon.$

□

Proof of Claim 28. We have

$$\left|\Big(\prod_{i=1}^k \mathbb{E}[X_i]\Big)\Big(\prod_{i=1}^k e^{\mathbb{E}[W_i]} - \prod_{i=1}^k e^{\mathbb{E}[W_i']}\Big) e^{\hat{W}}\right| = \Big|\prod_{i=1}^k \mathbb{E}[X_i]\Big| \cdot \Big|\prod_{i=1}^k e^{\mathbb{E}[W_i]} - \prod_{i=1}^k e^{\mathbb{E}[W_i']}\Big| \cdot \Big|e^{\hat{W}}\Big|$$

Since $|\sum_i \mathbb{E}[W_i] - \sum_i \mathbb{E}[W_i']| \le \sum_i |\mathbb{E}[W_i] - \mathbb{E}[W_i']| \le k\varepsilon$, by Claim 21,

$$\begin{aligned} \Big|\prod_{i=1}^k e^{\mathbb{E}[W_i]} - \prod_{i=1}^k e^{\mathbb{E}[W_i']}\Big| &= \Big|e^{\sum_i \mathbb{E}[W_i]} - e^{\sum_i \mathbb{E}[W_i']}\Big| \\ &\le \Big|e^{\sum_i \mathbb{E}[W_i]}\Big| \cdot O(k\varepsilon) \\ &= \Big|\prod_{i=1}^k e^{\mathbb{E}[W_i]}\Big| \cdot O(k\varepsilon), \end{aligned}$$

Therefore,

$$\begin{aligned} \Big|\prod_{i=1}^k \mathbb{E}[X_i]\Big| \cdot \Big|\prod_{i=1}^k e^{\mathbb{E}[W_i]} - \prod_{i=1}^k e^{\mathbb{E}[W_i']}\Big| \cdot \Big|e^{\hat{W}}\Big| &\le \Big|\prod_{i=1}^k \mathbb{E}[X_i]\Big| \cdot \Big|\prod_{i=1}^k e^{\mathbb{E}[W_i]}\Big| \cdot O(k\varepsilon) \cdot \Big|e^{\hat{W}}\Big| \\ &= \left|\Big(\prod_{i=1}^k \mathbb{E}[X_i] e^{\mathbb{E}[W_i]}\Big) e^{\hat{W}}\right| \cdot O(k\varepsilon) \\ &= \Big|\prod_{i=1}^k X_i\Big| \cdot O(k\varepsilon) \\ &\le O(k\varepsilon). \end{aligned}$$

□

Proof of Claim 29. We first rewrite $e^{\hat{W}} - e^{\hat{W}'}$ as a sum of 3 terms:

$$e^{\hat{W}} - e^{\hat{W}'} = \Big(e^{\hat{W}} - \sum_{j=0}^{d-1} \hat{W}^j/j!\Big) + \Big(\sum_{j=0}^{d-1} (\hat{W}^j - \hat{W'}^j)/j!\Big) + \Big(\sum_{j=0}^{d-1} \hat{W'}^j/j! - e^{\hat{W}'}\Big).$$

It suffices to bound above the expectation of each term multiplied by $\gamma := \prod_{i=1}^k \mathbb{E}[X_i] e^{\mathbb{E}[W_i']}$. We bound the first and last terms using Taylor's approximation (Lemma 19), and the second term using (ε, d)-closeness of the variables. We will show the following:

$$\mathbb{E}\left[\left|\gamma \cdot \Big(e^{\hat{W}'} - \sum_{j=0}^{d-1} \hat{W'}^j/j!\Big)\right|\right] \le 2^{O(d)} \left(\frac{\sigma(Z)\sqrt{d} + Bd}{d}\right)^d + (kB)^{O(d)} \varepsilon \tag{7}$$

$$\mathbb{E}\left[\left|\gamma \cdot \Big(e^{\hat{W}} - \sum_{j=0}^{d-1} \hat{W}^j/j!\Big)\right|\right] \le 2^{O(d)} \left(\frac{\sigma(Z)\sqrt{d} + Bd}{d}\right)^d \tag{8}$$

$$\left|\gamma \cdot \mathbb{E}\Big[\sum_{j=0}^{d-1} (\hat{W}^j - \hat{W'}^j)/j!\Big]\right| \le k^d \varepsilon. \tag{9}$$

For (7), by Lemma 19 we have

$$\left|\gamma \cdot \Big(e^{\hat{W}'} - \sum_{j=0}^{d-1} \hat{W'}^j/j!\Big)\right| \le |\gamma| \cdot O(1) \frac{|\hat{W}'|^d}{d!} \cdot \max\{1, e^{\Re(\hat{W}')}\}.$$

We now bound above $|\gamma \cdot \max\{1, e^{\Re(\hat{W}')}\}|$ by 1. We have

$$\begin{aligned}
|\gamma| &= \Big|\prod_{i=1}^k \mathbb{E}[X_i] e^{\mathbb{E}[W_i']}\Big| \\
&= \Big|\prod_{i=1}^k \mathbb{E}[X_i]\Big| \cdot \Big|e^{\mathbb{E}[\sum_i W_i']}\Big| \\
&\le \Big|\prod_{i=1}^k \mathbb{E}[X_i]\Big| \cdot \mathbb{E}[|e^{\sum_i W_i'}|] && \text{(Jensen's inequality, see Lemma 20)} \\
&= \mathbb{E}\left[\Big|\prod_{i=1}^k \mathbb{E}[X_i] \cdot e^{\sum_i W_i'}\Big|\right] \\
&= \mathbb{E}\left[\Big|\prod_{i=1}^k Y_i\Big|\right] \\
&\le 1.
\end{aligned}$$

Moreover,

$$|\gamma \cdot e^{\Re(\hat{W}')}| = \Big|\prod_{i=1}^k \mathbb{E}[X_i] e^{\mathbb{E}[W_i']}\Big| \cdot e^{\Re(\hat{W}')} = \Big|\prod_{i=1}^k \mathbb{E}[X_i] e^{\mathbb{E}[W_i']} e^{\hat{W}'}\Big| = \Big|\prod_{i=1}^k Y_i\Big| \le 1.$$

Hence, it suffices to bound above $\mathbb{E}[|\hat{W}'|^d]$. Note that the $\hat{W}_i'$'s are (ε, d)-close to the $\hat{W}_i$'s. So we bound above $|\hat{W}_i|$ and $\mathrm{Var}[\hat{W}_i]$ and then apply Lemma 47. First, since $|Z_i| \le B$, we have $|W_i| \le 2B$ because of Lemma 17, and so $|\hat{W}_i| \le |W_i| + |\mathbb{E}[W_i]| \le 4B$. Next, we have $\mathrm{Var}[\hat{W}_i] \le 4\,\mathrm{Var}[Z_i]$ because of Lemma 18, and so $\sigma(\hat{W}) \le 2\sigma(Z)$. Therefore, by Lemma 47,

$$\begin{aligned}
\mathbb{E}\left[\left|\Big(\prod_{i=1}^{k} \mathbb{E}[X_i] e^{\mathbb{E}[W_i']}\Big)\Big(e^{\hat{W}'} - \sum_{j=0}^{d-1} \hat{W'}^j/j!\Big)\right|\right] &\le O(1)\frac{\mathbb{E}[|\hat{W}'|^d]}{d!} \\
&\le 2^{O(d)}\left(\frac{\sigma(\hat{W})\sqrt{d} + 4Bd}{d}\right)^d + (kB)^{O(d)}\varepsilon \\
&\le 2^{O(d)}\left(\frac{\sigma(Z)\sqrt{d} + Bd}{d}\right)^d + (kB)^{O(d)}\varepsilon.
\end{aligned}$$

We prove Inequality (8) similarly. Note that

$$\begin{aligned}
\frac{|e^{\sum_i \mathbb{E}[W_i']}|}{|e^{\sum_i \mathbb{E}[W_i]}|} &= |e^{\sum_i \mathbb{E}[W_i'] - \sum_i \mathbb{E}[W_i]}| \\
&\le e^{|\sum_i \mathbb{E}[W_i'] - \sum_i \mathbb{E}[W_i]|} \\
&\le e^{\sum_i |\mathbb{E}[W_i'] - \mathbb{E}[W_i]|} \\
&\le e^{k\varepsilon} \\
&\le O(1),
\end{aligned}$$

because $\varepsilon < 1/k$, otherwise the conclusion is trivial. Hence,

$$\left|\Big(\prod_{i=1}^{k} \mathbb{E}[X_i] e^{\mathbb{E}[W_i']}\Big)\Big(e^{\hat{W}} - \sum_{j=0}^{d-1} \hat{W}^j/j!\Big)\right| \le \left|\Big(\prod_{i=1}^{k} \mathbb{E}[X_i] e^{\mathbb{E}[W_i]}\Big)\Big(e^{\hat{W}} - \sum_{j=0}^{d-1} \hat{W}^j/j!\Big)\right| \cdot O(1).$$

Therefore, it follows by Inequality (1) by considering $\varepsilon = 0$ that

$$\mathbb{E}\left[\left|\Big(\prod_{i=1}^{k} \mathbb{E}[X_i] e^{\mathbb{E}[W_i']}\Big)\Big(e^{\hat{W}} - \sum_{j=0}^{d-1} \hat{W}^j/j!\Big)\right|\right] \le 2^{O(d)}\left(\frac{\sigma(Z)\sqrt{d} + Bd}{d}\right)^d.$$

Finally we prove Inequality (9). By linearity of expectation,

$$\mathbb{E}\Big[\sum_{j=0}^{d-1} (\hat{W}^j - \hat{W'}^j)/j!\Big] = \sum_{j=0}^{d-1} (\mathbb{E}[\hat{W}^j] - \mathbb{E}[\hat{W'}^j])/j!\,.$$

Note that $\hat{W}^j = (\sum_i \hat{W}_i)^j$ can be written as a sum of k^j terms where each term is a product of at most $j \le d$ different W_i's. Moreover, we have $|W_i| \le 2B \le 1$ for each i because of

Lemma 17. So we have $|\mathbb{E}[\hat{W}^j] - \mathbb{E}[\hat{W'}^j]| \leq k^j \varepsilon$. Hence,

$$\begin{aligned}
\left| \mathbb{E}\left[\sum_{j=0}^{d-1} (\hat{W}^j - \hat{W'}^j)/j! \right] \right| &\leq \sum_{j=0}^{d-1} |\mathbb{E}[\hat{W}^j] - \mathbb{E}[\hat{W'}^j]| \\
&\leq \sum_{j=0}^{d-1} k^j \varepsilon \\
&\leq k^d \varepsilon.
\end{aligned}$$

Recall that $|\gamma| \leq 1$, this concludes (9). □

4.2.3 Assuming the expectations are large

We now prove the main lemma assuming the expectation of the X_i are far from zero.

Lemma 30. *Let $X_1, X_2, \ldots, X_k$ be k independent random variables over $\mathbb{C}_{\leq 1}$, with $\min_{z \in \mathrm{Supp}(X_i)} \Pr[X_i = z] \geq 2^{-n}$. Let $Y_1, Y_2, \ldots, Y_k$ be k random variables over $\mathbb{C}_{\leq 1}$ that are $(\varepsilon, 9d)$-close to $X_1, \ldots, X_k$. Assume $|\mathbb{E}[X_i]| \geq (\sigma(X)^2/d)^{1/6}$ for each i. We have*

$$\left| \mathbb{E}\left[\prod_{i=1}^{k} X_i \right] - \mathbb{E}\left[\prod_{i=1}^{k} Y_i \right] \right| \leq 2^{O(d)} \left(\frac{\sigma(X)^2}{d} \right)^d + (k 2^n)^{O(d)} \varepsilon.$$

Proof of Lemma 30. We will assume $\sigma(X)^2/d$ is less than a sufficiently small constant and $\varepsilon \leq (k2^n)^{-Cd}$ for a sufficiently large C; otherwise the R.H.S. of the inequality is greater than 2 and there is nothing to prove.

For each $i \in \{1, 2, \ldots, k\}$, we define a new function $\mathrm{rd}_i \colon \mathbb{C}_{\leq 1} \to \mathbb{C}_{\leq 1}$ that will be used to round the variables X_i and Y_i. We define rd_i as

$$\mathrm{rd}_i(z) := \begin{cases} z & \text{if } |z - \mathbb{E}[X_i]| \leq (\sigma(X)^2/d)^{1/3} \\ \mathbb{E}[X_i] & \text{otherwise.} \end{cases}$$

Let $\tilde{X}_i = \mathrm{rd}_i(X_i)$ and $\tilde{Y}_i = \mathrm{rd}_i(Y_i)$. We will write both $\prod_i X_i$ and $\prod_i Y_i$ as

$$\prod_{i=1}^{k} X_i = \prod_{i=1}^{k} (X_i - \tilde{X}_i + \tilde{X}_i) = \sum_{S \subseteq \{1,2,\ldots,k\}} \prod_{i \in S} (X_i - \tilde{X}_i) \prod_{i \notin S} \tilde{X}_i,$$

and

$$\prod_{i=1}^{k} Y_i = \prod_{i=1}^{k} (Y_i - \tilde{Y}_i + \tilde{Y}_i) = \sum_{S \subseteq \{1,2,\ldots,k\}} \prod_{i \in S} (Y_i - \tilde{Y}_i) \prod_{i \notin S} \tilde{Y}_i.$$

Let $m = 3d$. Define

$$P_m(z_1, z_2, \ldots, z_k) = \sum_{|S| < m} \prod_{i \in S} (z_i - \mathrm{rd}_i(z_i)) \prod_{i \notin S} \mathrm{rd}_i(z_i).$$

We will show that P_m is a good approximation of the product in expectation over both X_i's and Y_i's and then show that the expectations of P_m under X_i's and Y_i's are close.

We will use the following inequalities repeatedly.

Claim 31. $\Pr[\tilde{X}_i \neq X_i] \leq \mathrm{Var}[X_i] v^{-2/3} \leq v^{1/3}$. *In particular,* $\sum_i \Pr[\tilde{X}_i \neq X_i] \leq (d\sigma)^{2/3}$.

Proof. The first inequality follows from Chebyshev's inequality and second follows from the assumption $\mathrm{Var}[X_i] \leq v$. The last sentence is implied by the first inequality. □

Claim 32. $\left|\mathbb{E}\left[\prod_i Y_i - P_m(Y_1, \ldots, Y_k)\right]\right| \leq 2^{O(d)} v^d + k^{O(d)}\varepsilon$.

Proof. Consider the product $\prod_{i \in S}(Y_i - \tilde{Y}_i)$. Let N' be the number of $i \in \{0, 1, 2, \ldots, k\}$ such that $\tilde{Y}_i \neq Y_i$. If $N' < m$ then any set S of size at least m must contain an i such that $\tilde{Y}_i = Y_i$. In this case the product is 0 and thus

$$\prod_i Y_i - P_m(Y_1, \ldots, Y_k) = \sum_{|S| \geq m} \prod_{i \in S}(Y_i - \tilde{Y}_i) \prod_{i \notin S} \tilde{Y}_i = 0.$$

So,

$$\begin{aligned}
\left|\mathbb{E}\left[\prod_i Y_i - P_m(Y_1, \ldots, Y_k)\right]\right| &= \left|\mathbb{E}\left[\mathbb{1}(N' \geq m) \cdot \left(\prod_i Y_i - P_m(Y_1, \ldots, Y_k)\right)\right]\right| \\
&\leq \mathbb{E}\left[\mathbb{1}(N' \geq m) \cdot \left(\left|\prod_i Y_i\right| + |P_m(Y_1, \ldots, Y_k)|\right)\right] \\
&= \mathbb{E}\left[\mathbb{1}(N' \geq m) \cdot \left|\prod_i Y_i\right|\right] + \mathbb{E}\left[\mathbb{1}(N' \geq m) \cdot |P_m(Y_1, \ldots, Y_k)|\right].
\end{aligned}$$

If $N' \geq m$ then there can be at most $\sum_{\ell=0}^{m-1} \binom{N'}{\ell} \leq \sum_{\ell=0}^{m-1} \binom{N'}{m}\binom{m}{\ell} \leq 2^m \binom{N'}{m}$ subsets in the sum in P_m for which the product is nonzero, and each such product can be at most 2^m because $|S| < m$. Thus,

$$\begin{aligned}
\mathbb{1}(N' \geq m) \cdot \left|P_m(Y_1, \ldots, Y_k)\right| &\leq \mathbb{1}(N' \geq m) \cdot 2^m \sum_{\ell=0}^{m-1} \binom{N'}{\ell} \\
&\leq \mathbb{1}(N' \geq m) \cdot 2^m \cdot 2^m \binom{N'}{m} \\
&\leq 2^{2m} \binom{N'}{m}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}(N' \geq m) \cdot \left(\left|\prod_i Y_i\right| + |P_m(Y_1, \ldots, Y_k)|\right)\right] &\leq \mathbb{E}\left[\mathbb{1}(N' \geq m) \cdot \left|\prod_i Y_i\right|\right] + 2^{2m}\, \mathbb{E}\left[\binom{N'}{m}\right] \\
&\leq \mathbb{E}[\mathbb{1}(N' \geq m)] + 2^{2m}\, \mathbb{E}\left[\binom{N'}{m}\right].
\end{aligned}$$

We will show the following

Claim 33. $\Pr[N' \geq m] \leq \mathbb{E}\left[\binom{N'}{m}\right] \leq v^d + k^{O(d)}\varepsilon$.

Assuming the claim it follows that

$$\begin{aligned}\left|\mathbb{E}\left[\mathbb{1}(N' \geq m) \cdot \Big(\prod_i Y_i - P_m(Y_1, \dots, Y_k)\Big)\right]\right| &\leq \mathbb{E}[\mathbb{1}(N' \geq m)] + 2^{2m}\,\mathbb{E}\left[\binom{N'}{m}\right] \\ &\leq (1 + 2^{6d})((2v)^d + k^{O(d)}\varepsilon) \qquad (m = 6d) \\ &\leq 2^{O(d)} v^d + k^{O(d)}\varepsilon. \qquad \square\end{aligned}$$

We now prove Claim 33.

Proof of Claim 33. The first inequality is clear.

$$\begin{aligned}\mathbb{E}\left[\binom{N'}{m}\right] &\leq \sum_{|S|=m} \Pr[\wedge_{i \in S} Y_i \neq \tilde{Y}_i] \\ &\leq \sum_{|S|=m} \Big(\prod_{i \in S} \Pr[X_i \neq \tilde{X}_i] + \varepsilon\Big) \qquad \text{(each } Y_i \text{ is } \varepsilon\text{-close to } X_i) \\ &\leq \sum_{|S|=m} \prod_{i \in S} \Pr[X_i \neq \tilde{X}_i] + k^m \varepsilon \\ &\leq \left(\frac{e \sum_{i=1}^k \Pr[X_i \neq \tilde{X}_i]}{m}\right)^m + k^m \varepsilon \qquad \text{(Maclaurin's inequality)} \\ &\leq \left(\frac{e(d \cdot \sigma(X))^{2/3}}{3d}\right)^{3d} + k^m \varepsilon \qquad \text{(Claim 31)} \\ &\leq v^d + k^{O(d)}\varepsilon. \qquad \square\end{aligned}$$

Now, we show that $P_m(Y_1, \dots, Y_k)$ is close to $P_m(X_1, \dots, X_k)$ in expectation.

Claim 34. $|\mathbb{E}[P_m(X_1, \dots, X_k)] - \mathbb{E}[P_m(Y_1, \dots, Y_k)]| \leq 2^{O(d)} v^d + O(k)^{3d}\varepsilon$.

Proof. The difference between $P_m(X_1, \dots, X_k)$ and $P_m(Y_1, \dots, Y_k)$ equals

$$P_m(X_1, \dots, X_k) - P_m(Y_1, \dots, Y_k) = \sum_{|S|<m} \Big(\prod_{i \in S}(X_i - \tilde{X}_i) \prod_{i \notin S} \tilde{X}_i - \prod_{i \in S}(Y_i - \tilde{Y}_i) \prod_{i \notin S} \tilde{Y}_i\Big).$$

We can rewrite the R.H.S. as

$$\sum_{|S|<m} \left(\Big(\prod_{i \in S}(X_i - \tilde{X}_i) - \prod_{i \in S}(Y_i - \tilde{Y}_i)\Big) \prod_{i \notin S} \tilde{X}_i + \prod_{i \in S}(Y_i - \tilde{Y}_i)\Big(\prod_{i \notin S} \tilde{X}_i - \prod_{i \notin S} \tilde{Y}_i\Big)\right).$$

It suffices to show that

$$\left|\mathbb{E}\left[\sum_{|S|<m}\Big(\prod_{i\in S}(X_i-\tilde{X}_i)-\prod_{i\in S}(Y_i-\tilde{Y}_i)\Big)\prod_{i\notin S}\tilde{X}_i\right]\right|\le k^{O(d)}\varepsilon \tag{10}$$

$$\left|\mathbb{E}\left[\sum_{|S|<m}\prod_{i\in S}(Y_i-\tilde{Y}_i)\Big(\prod_{i\notin S}\tilde{X}_i-\prod_{i\notin S}\tilde{Y}_i\Big)\right]\right|\le 2^{O(d)}v^d+(k2^n)^{O(d)}\varepsilon. \tag{11}$$

We first prove Inequality (10). Because the X_i's are independent, the L.H.S. of the inequality equals

$$\begin{aligned}
&\left|\sum_{|S|<m}\left(\mathbb{E}\Big[\prod_{i\in S}(X_i-\tilde{X}_i)\Big]-\mathbb{E}\Big[\prod_{i\in S}(Y_i-\tilde{Y}_i)\Big]\right)\mathbb{E}\Big[\prod_{i\notin S}\tilde{X}_i\Big]\right|\\
&\le\sum_{\ell=1}^{m-1}\sum_{|S|=\ell}\left|\mathbb{E}\Big[\prod_{i\in S}(X_i-\tilde{X}_i)\Big]-\mathbb{E}\Big[\prod_{i\in S}(Y_i-\tilde{Y}_i)\Big]\right|\cdot\left|\mathbb{E}\Big[\prod_{i\notin S}\tilde{X}_i\Big]\right|\\
&\le\sum_{\ell=1}^{m-1}\sum_{|S|=\ell}\left|\mathbb{E}\Big[\prod_{i\in S}(X_i-\tilde{X}_i)\Big]-\mathbb{E}\Big[\prod_{i\in S}(Y_i-\tilde{Y}_i)\Big]\right|\\
&\le\sum_{\ell=1}^{m-1}\sum_{|S|=\ell}2\cdot 2^{\ell}\varepsilon\\
&\le\sum_{\ell=1}^{m-1}k^{\ell}\cdot 2\cdot 2^{\ell}\varepsilon\\
&\le 2(2k)^m\varepsilon\\
&=k^{O(d)}\varepsilon.
\end{aligned}$$

To see the third inequality, note that $|z-\mathrm{rd}_i(z)|\le 2$, and so $|\prod_{i\in S}(z_i-\mathrm{rd}_i(z_i))|\le 2^{|S|}$. So we can apply Claim 22 to bound above the absolute difference by $2\cdot 2^{|S|}\varepsilon$.

Now we prove Inequality (11). As $|S|\le m=3d$ and Y_i's are $(\varepsilon,9d)$-close to X_i's, conditioned on the values of $\tilde{X}_i$ for which $i\in S$, by Claim 24, the remaining $\tilde{Y}_i$'s for which $i\notin S$ are still $(2^{O(m\cdot n)}\varepsilon,6d)$-close to the corresponding $\tilde{X}_i$'s. (Recall that we can assume $\varepsilon=(k2^n)^{-Cd}$ for a sufficiently large C.) We will apply Lemma 26 to them.

Define Z_i,Z'_i such that $\tilde{X}_i=\mathbb{E}[\tilde{X}_i](1+Z_i)$ and $\tilde{Y}_i=\mathbb{E}[\tilde{X}_i](1+Z'_i)$. To apply Lemma 26, we need the following two claims to bound above $|Z_i|,|Z'_i|$ and $\sigma(Z)^2$. We defer their proofs to the end.

Claim 35. *Let $B=4v^{1/6}$. Then $|Z_i|\le B$ and $|Z'_i|\le B$.*

Claim 36. $\sigma(Z)^2\le 4\sigma(X)^2v^{-1/3}$.

Therefore, by Lemma 26 with $\varepsilon' = 2^{O(m\cdot n)}\varepsilon$ and $B = 4(\sigma(X)^2/d)^{1/6} \leq 1/2$ (Recall that we can assume $\sigma(X)^2/d$ less than a sufficiently small constant),

$$\left|\mathbb{E}\left[\sum_{|S|<m}\prod_{i\in S}(Y_i - \tilde{Y}_i)\Big(\prod_{i\notin S}\tilde{X}_i - \prod_{i\notin S}\tilde{Y}_i\Big)\right]\right| \leq \sum_{|S|<m}\mathbb{E}\left[\left|\prod_{i\in S}(Y_i - \tilde{Y}_i)\right|\right]\cdot M,$$

where

$$\begin{aligned}
M &\leq 2^{O(d)}\left(\frac{\sigma(Z)\sqrt{d} + dB}{d}\right)^{6d} + (Bk)^{O(d)}\varepsilon' \\
&\leq 2^{O(d)}\left(\frac{\sigma(X)(\sigma(X)/\sqrt{d})^{-1/3}}{\sqrt{d}} + B\right)^{6d} + (Bk2^n)^{O(d)}\varepsilon \\
&\leq 2^{O(d)}\left(\frac{\sigma(X)(\sigma(X)/\sqrt{d})^{-1/3}}{\sqrt{d}} + 4v^{1/6}\right)^{6d} + (k2^n)^{O(d)}\varepsilon \\
&= 2^{O(d)}\left(v^{1/3} + v^{1/6}\right)^{6d} + (k2^n)^{O(d)}\varepsilon \\
&= 2^{O(d)}v^d + (k2^n)^{O(d)}\varepsilon.
\end{aligned}$$

We now bound above $\mathbb{E}[|\prod_{i\in S}(Y_i - \tilde{Y}_i)|]$. Note that $|\prod_{i\in S}(z_i - \mathrm{rd}_i(z_i))| \leq 2^{|S|}$. Hence by Claim 22,

$$\mathbb{E}\left[\left|\prod_{i\in S}(Y_i - \tilde{Y}_i)\right|\right] \leq \mathbb{E}\left[\left|\prod_{i\in S}(X_i - \tilde{X}_i)\right|\right] + 2^{|S|}\varepsilon.$$

Let N be the number of $i \in \{0, 1, \ldots, k\}$ such that $\tilde{X}_i \neq X_i$. Note that

$$\begin{aligned}
\sum_{|S|<m}\mathbb{E}\left[\left|\prod_{i\in S}(X_i - \tilde{X}_i)\right|\right] &\leq \sum_{\ell=0}^{m-1}\left(2^\ell\,\mathbb{E}\left[\binom{N}{\ell}\right]\right) \\
&\leq 2^m\,\mathbb{E}[2^N] \\
&= 2^m\prod_{i=1}^{k}\Big(1 + \Pr[X_i \neq \tilde{X}_i]\Big) \\
&\leq 2^m e^{\sum_i \Pr[X_i \neq \tilde{X}'_i]} \\
&\leq 2^m e^{(d\,\sigma(X))^{2/3}} \\
&\leq 2^{O(d)},
\end{aligned}$$

where the last inequality is because $\sigma(X)^2/d \leq 1$ and so $\sigma(X)^{2/3} \leq d^{1/3}$. Therefore,

$$\sum_{|S|<m}\mathbb{E}\left[\left|\prod_{i\in S}(Y_i - \tilde{Y}_i)\right|\right] \leq 2^{O(d)} + \sum_{|S|<m}2^{|S|}\varepsilon \leq 2^{O(d)} + (2k)^m\varepsilon \leq 2^{O(d)},$$

where the last inequality is because $\varepsilon \leq k^{-Cd}$ for a sufficiently large C. So altogether the bound is $2^{O(d)} \cdot M$ as desired. □

□

We now prove Claim 35 and 36. By Claim 31, $|\mathbb{E}[X_i] - \mathbb{E}[\tilde{X}_i]| \leq (\sigma(X)^2/d)^{1/3}$. Also by assumption, $|\mathbb{E}[X_i]| \geq (\sigma(X)^2/d)^{1/6}$. So, we have $|\mathbb{E}[\tilde{X}_i]| \geq |\mathbb{E}[X_i]|/2 \geq (\sigma(X)^2/d)^{1/6}/2$.

Proof of Claim 35. As $|\mathbb{E}[\tilde{X}_i]| \geq v^{1/6}/2$, we have

$$\begin{aligned}
|\tilde{Z}_i| &= \frac{|\tilde{X}_i - \mathbb{E}[\tilde{X}_i]|}{|\mathbb{E}[\tilde{X}_i]|} \\
&\leq \frac{|\tilde{X}_i - \mathbb{E}[X_i]| + |\mathbb{E}[\tilde{X}_i] - \mathbb{E}[X_i]|}{|\mathbb{E}[\tilde{X}_i]|} \\
&\leq 4v^{1/3}/v^{1/6} \\
&\leq 4v^{1/6},
\end{aligned}$$

and the same argument holds for $|\tilde{Z}'_i|$ because $|\tilde{Y}_i - \mathbb{E}[X_i]| \leq v^{1/3}$. □

Proof of Claim 36. Since $z^* = \mathbb{E}[Z]$ is the minimizer of $\mathbb{E}[|Z - z|^2]$, we have

$$\begin{aligned}
\mathrm{Var}[\tilde{X}_i] &= \mathbb{E}[|\tilde{X}_i - \mathbb{E}[\tilde{X}_i]|^2] \\
&\leq \mathbb{E}[|\tilde{X}_i - \mathbb{E}[X_i]|^2] \\
&\leq \mathbb{E}[|X_i - \mathbb{E}[X_i]|^2] && (\tilde{X}_i = \mathrm{rd}_i(X_i)) \\
&= \mathrm{Var}[X_i].
\end{aligned}$$

Therefore, $\mathrm{Var}[\tilde{Z}_i] = \mathrm{Var}[\tilde{X}_i]/|\mathbb{E}[\tilde{X}_i]|^2 \leq 4\,\mathrm{Var}[X_i]v^{-1/3}$ and thus $\sum_i \mathrm{Var}[\tilde{Z}_i] \leq 4\sigma(X)^2 v^{-1/3}$. □

4.2.4 The general case

Proof of Lemma 12. We will again assume $\sigma(X)^2/d$ is less than a sufficiently small constant and $\varepsilon \leq (k2^n)^{-Cd}$ for a sufficiently large constant C. We first assume $\mathrm{Var}[X_j] \leq \sigma(X)^2/d$ for all j and prove the lemma when the Y_i's are $(\varepsilon, 15d)$-close to the X_i's. Later we will handle the general case.

Let m be the number of i such that $|\mathbb{E}[X_i]| \leq v^{1/6}$.

If $m \leq 6d$, let J be the set of indices for which $|\mathbb{E}[X_i]| \leq v^{1/6}$. We can write

$$\prod_i X_i - \prod_i Y_i = \left(\prod_{j\in J} X_j - \prod_{j\in J} Y_j\right) \prod_{j\notin J} X_j + \prod_{j\in J} Y_j \left(\prod_{j\notin J} X_j - \prod_{j\notin J} Y_j\right).$$

It suffices to show that

$$\left|\mathbb{E}\left[\Big(\prod_{j\in J}X_j-\prod_{j\in J}Y_j\Big)\prod_{j\notin J}X_j\right]\right|\le\varepsilon \tag{12}$$

$$\left|\mathbb{E}\left[\prod_{j\in J}Y_j\left(\prod_{j\notin J}X_j-\prod_{j\notin J}Y_j\right)\right]\right|\le 2^{O(d)}v^d+(k2^n)^{O(d)}\varepsilon. \tag{13}$$

We first show Inequality (12). Since the X_i's are independent, the L.H.S. of (12) is

$$\begin{aligned}\left|\left(\mathbb{E}\Big[\prod_{j\in J}X_j\Big]-\mathbb{E}\Big[\prod_{j\in J}Y_j\Big]\right)\mathbb{E}\Big[\prod_{j\notin J}X_j\Big]\right| &\le \left|\mathbb{E}\Big[\prod_{j\in J}X_j\Big]-\mathbb{E}\Big[\prod_{j\in J}Y_j\Big]\right|\\ &\le\varepsilon.\end{aligned}$$

To prove Inequality (13), note that conditioned on the values of the Y_i's for which $i\in J$, by Claim 24, the rest of the Y_i's are still $(2^{O(dn)}\varepsilon,9d)$-close to the corresponding X_i's with $|\mathbb{E}[X_i]|\ge v^{1/6}$. (Recall that we can assume $\varepsilon=(k2^n)^{-Cd}$ for a sufficiently large C.) So the bound follows from Lemma 30.

If $m\ge 6d$, then note that

$$\left|\mathbb{E}\Big[\prod_{i=1}^k X_i\Big]\right|=\prod_{i=1}^k|\mathbb{E}[X_i]|\le v^{m/6}\le v^d.$$

So it suffices to show that

$$\left|\mathbb{E}\Big[\prod_{i=1}^k Y_i\Big]\right|\le 2^{O(d)}v^d+k^{O(d)}\varepsilon.$$

Consider the event E that at least $3d$ of the Y_i for $i\in J$ have absolute value less than $2v^{1/6}$. Then we know that

$$\Big|\prod_{i=1}^k Y_i\Big|\le 2^{3d}\cdot v^{d/2}.$$

We will show that E happens except with probability at most $v^{2d}+k^{3d}\varepsilon$. Let $N\in\{0,1,2,\dots,m\}$ be the number of $i\in J$ such that $|Y_i|\ge 2v^{1/6}$. Note that

$$\begin{aligned}\Pr[N\ge 3d] &\le \sum_{S\subseteq J:|S|=3d}\Pr\left[\bigwedge_{i\in S}\big(|Y_i|\ge 2v^{1/6}\big)\right]\\ &\le \sum_{S\subseteq J:|S|=3d}\prod_{i\in S}\Pr\left[|X_i|\ge 2v^{1/6}\right]+k^{3d}\varepsilon.\end{aligned}$$

By Chebyshev's inequality,

$$\Pr[|X_i|\ge 2v^{1/6}]\le\Pr[|X_i-\mathbb{E}[X_i]|\ge v^{1/6}]\le\mathrm{Var}[X_i]v^{-1/3}.$$

Hence, by Maclaurin's inequality,

$$\begin{aligned}\sum_{S\subseteq J:|S|=3d}\prod_{i\in S}\Pr\left[|X_i|\geq 2v^{1/6}\right] &\leq \left(\frac{e\sum_{i=1}^m \Pr[|X_i|\geq 2v^{1/6}]}{3d}\right)^{3d}\\ &\leq \left(\frac{e\sum_{i=1}^m \mathrm{Var}[X_i]v^{-1/3}}{3d}\right)^{3d}\\ &\leq \left(\frac{e\sigma(X)^2v^{-1/3}}{3d}\right)^{3d}\\ &\leq v^{2d}.\end{aligned}$$

So,

$$\Pr[N\geq 3d]\leq v^{2d}+k^{3d}\varepsilon.$$

Therefore,

$$\begin{aligned}\left|\mathbb{E}\left[\prod_i Y_i\right]\right| &\leq 2^{3d}v^{d/2}+v^{2d}+k^{3d}\varepsilon\\ &\leq 2^{O(d)}v^{d/2}+k^{O(d)}\varepsilon.\end{aligned}$$

□

5 Improved bound for bounded independence plus noise fools products

In this section we prove Theorem 10, which improves the error bound in Theorem 8 from $2^{-\Omega(d)}$ to $n^{-\Omega(d)}$. Its proof requires developing a few additional technical tools. We first outline the high-level idea on how to obtain the improvement.

For simplicity, we will assume $d = O(1)$ and show how to obtain an error bound of $n^{-\Omega(1)}$. Recall in the proof of Theorem 8 (see also Table 1) that we used a win-win argument on the total-variance: we applied two different arguments depending on whether the total-variance of a product test f is above or below a certain threshold. Suppose now the total-variance of f is guaranteed to lie outside the interval $[n^{-0.1}, n^{0.1}]$. Then applying the same arguments as before would already give us an error of $n^{-\Omega(1)}$. So it suffices to handle the additional case, where the total-variance is in the range of $[n^{-0.1}, n^{0.1}]$. Our goal is to use noise to reduce the total-variance down to $n^{-0.1}$, which can then be handled by the low total-variance argument. To achieve this, as a first step we will handle the functions f_i with variances above and below $n^{-0.6}$ separately, and show that $O(n)$-wise independence plus noise fools the product of the f_i in each case.

For the former, note that since the total-variance is $\leq n^{0.1}$, there can be at most $n^{0.7}$ functions with variances above $n^{-0.6}$. In this case we can simply apply the result in [HLV17] (Theorem 7). To prove the latter case, we use noise to reduce the variance of each function. Specifically, we use the hypercontractivity theorem to show that applying the noise operator to a function reduces its variance from σ^2 to $(\sigma^2)^{(4/3)}$. This is proved in Section 5.1

below. Hence, on average over the noise, the variance σ_i^2 of each f_i is reduced to at most $(n^{-0.6})^{1/3}\sigma_i^2$, and so the total-variance of the f_i is at most $(n^{-0.6})^{1/3} \cdot n^{0.1} = n^{-0.1}$ and we can argue as before. To combine the two cases, we prove a new XOR Lemma for bounded independent distributions, inspired by a similar lemma for small-bias distributions which is proved in [GMR+12], and the theorem follows.

5.1 Noise reduces variance of bounded complex-valued functions

In this section, we show that on average, noise reduces the variance of bounded complex-valued functions. We will use the hypercontractivity theorem for complex-valued functions (cf. [Hat14, Theorem 6.1.8]).

Let $f\colon \{0,1\}^n \to \mathbb{C}$ be any function. For every $\rho \in [0,1]$, define the noise operator T_ρ to be $T_\rho f(x) := \mathbb{E}_N[f(x+N)]$, where N sets each bit to uniform independently with probability $1-\rho$ and 0 otherwise.

Theorem 37 (Hypercontractivity Theorem). *Let $q \in [2,\infty)$. Then for any $\rho \in [0, \sqrt{1/(q-1)}]$,*

$$\mathbb{E}\big[|T_\rho f(x)|^q\big]^{1/q} \le |\mathbb{E}[f(x)^2]|^{1/2}.$$

We will use the following well-known corollary.

Corollary 38. *Let $f\colon \{0,1\}^n \to \mathbb{C}$. Then*

$$\mathbb{E}\big[|T_\rho f(x)|^2\big] \le \mathbb{E}\big[|f(x)|^{1+\rho^2}\big]^{\frac{2}{1+\rho^2}}.$$

Proof.

$$\begin{aligned}
\mathbb{E}\big[|T_\rho f(x)|^2\big] &= \mathop{\mathbb{E}}_x\big[\mathop{\mathbb{E}}_{N,N'}[f(x+N)\overline{f(x+N')}]\big] \\
&= \mathop{\mathbb{E}}_x\big[\mathop{\mathbb{E}}_{N,N'}[f(x)\overline{f(x+N+N')}]\big] \\
&= \mathop{\mathbb{E}}_x\big[f(x)\mathop{\mathbb{E}}_{N,N'}\overline{[f(x+N+N')]}\big] \\
&= \mathop{\mathbb{E}}_x\big[f(x)\overline{T_\rho T_\rho f(x)}\big] \\
&\le \mathbb{E}\big[|f(x)|^{1+\rho^2}\big]^{\frac{1}{1+\rho^2}}\,\mathbb{E}\big[|T_\rho T_\rho f(x)|^{1+\frac{1}{\rho^2}}\big]^{\frac{1}{1+1/\rho^2}} \\
&\le \mathbb{E}[|f(x)|^{1+\rho^2}]^{\frac{1}{1+\rho^2}}\,\mathbb{E}[|T_\rho f(x)|^2]^{1/2}.
\end{aligned}$$

The first inequality follows from Hölder's inequality because $\frac{1}{1+\rho^2} + \frac{1}{1+1/\rho^2} = 1$, and the second inequality follows from Theorem 37 with $q = 1 + 1/\rho^2$. □

Let T be a distribution over $\{0,1\}^m$ that sets each bit independently to 1 with probability $1-\rho$ and 0 otherwise.

Claim 39. $\mathbb{E}_{T,U}\big[|\mathbb{E}_{U'}[U + T \wedge U']|^2\big] = \mathbb{E}\big[|T_{\sqrt{\rho}} f(x)|^2\big]$.

Proof.

$$\begin{aligned}\mathop{\mathbb{E}}_{T,U}\big[|\mathop{\mathbb{E}}_{U'}[U+T\wedge U']|^2\big] &= \mathop{\mathbb{E}}_{T}\Big[\sum_{\alpha,\alpha'}\hat{f}_\alpha\overline{\hat{f}_{\alpha'}}\mathop{\mathbb{E}}_{U}[\chi_{\alpha+\alpha'}(U)]\mathop{\mathbb{E}}_{U'}[\chi_\alpha(T\wedge U')]\mathop{\mathbb{E}}_{U''}[\chi_{\alpha'}(T\wedge U'')]\Big]\\ &= \sum_\alpha|\hat{f}_\alpha|^2\mathop{\mathbb{E}}_{T,U',U''}\big[\chi_\alpha\big(T\wedge(U'+U'')\big)\big]\\ &= \sum_\alpha|\hat{f}_\alpha|^2\rho^{|\alpha|} = \mathbb{E}\big[|T_{\sqrt{\rho}}f(x)|^2\big],\end{aligned}$$

where the last inequality follows from Parseval's identity because the Fourier expansion of $T_\rho f(x)$ is $\sum_\alpha \hat{f}_\alpha\rho^{|\alpha|}\chi_\alpha(x)$. □

We are now ready to prove that noise reduces the variance of a function. The main idea is to translate the function to a point close to its mean so that its variance is close to its second moment, and then apply Corollary 38 to it.

Lemma 40. *Let $f\colon\{0,1\}^n\to\mathbb{C}_{\le 1}$ be any function. Let $\delta := \min\{|f(x)-f(x')| : f(x)\ne f(x')\}$. Then*

$$\mathop{\mathbb{E}}_{T}\Big[\mathop{\mathrm{Var}}_{x}\big[\mathop{\mathbb{E}}_{U}[f(x+T\wedge U)]\big]\Big] \le 4\left(\frac{2\,\mathrm{Var}[f]}{\delta^2}\right)^{\frac{2}{1+\rho}}.$$

Proof. We can assume $\mathrm{Var}[f]\le\delta^2/2$; otherwise the conclusion is trivial. Let S be the support of f. For every $y\in S$, let $p_y := \Pr[f(x)=y]$. Let $\mu=\mathbb{E}[f]$ and $\sigma^2=\mathrm{Var}[f]$. Since $\sigma^2=\mathbb{E}[|f(x)-\mu|^2]$, there is a point $z\in S$ such that $|z-\mu|^2\le\sigma^2$. We have

$$\sigma^2=\sum_{y\in S}p_y|y-\mu|^2 \ge \sum_{y\in S:y\ne z}p_y|y-\mu|^2 \ge \min_{y\in S:y\ne z}|y-\mu|^2\Big(\sum_{y\in S:y\ne z}p_y\Big).$$

Define $g(x) := \frac{f(x)-z}{2}$. We have for every t,

$$\mathop{\mathrm{Var}}_{x}\big[\mathop{\mathbb{E}}_{U}[f(x+t\wedge U)]\big] = 4\mathop{\mathrm{Var}}_{x}\big[\mathop{\mathbb{E}}_{U}[g(x+t\wedge U)]\big] \le 4\mathop{\mathbb{E}}_{x}\big[\big|\mathop{\mathbb{E}}_{U}[g(x+t\wedge U)]\big|^2\big].$$

By Corollary 38,

$$\mathbb{E}\big[|T_\rho g|^2\big] \le \mathbb{E}\big[|g|^{1+\rho^2}\big]^{\frac{2}{1+\rho^2}} = \left(\sum_{y\in S:y\ne z}p_y\left|\frac{y-z}{2}\right|^{1+\rho^2}\right)^{\frac{2}{1+\rho^2}} \le \Big(\sum_{y\in S:y\ne z}p_y\Big)^{\frac{2}{1+\rho^2}}$$

because $|y-y'|\le 2$ for every $y,y'\in\mathbb{C}_{\le 1}$. So by Claim 39, we have

$$\mathop{\mathbb{E}}_{T,x}\big[\big|\mathop{\mathbb{E}}_{U}[g(x+T\wedge U)]\big|^2\big] = \mathbb{E}\big[|Tg|^2\big] \le \Big(\sum_{y\in S:y\ne z}p_y\Big)^{\frac{2}{1+\rho}}.$$

It follows from above that

$$\mathop{\mathbb{E}}_{T}\Big[\mathop{\mathrm{Var}}_{x}\big[\mathop{\mathbb{E}}_{U}[f(x+T\wedge U)]\big]\Big] \le 4\mathop{\mathbb{E}}_{T,x}\big[\big|\mathop{\mathbb{E}}_{U}[g(x+T\wedge U)]\big|^2\big] \le 4\Big(\sum_{y\in S:y\ne z}p_y\Big)^{\frac{2}{1+\rho}} \le 4\left(\frac{\mathrm{Var}[f]}{\min_{y\in S:y\ne z}|y-\mu|^2}\right)^{\frac{2}{1+\rho}}.$$

Now we bound below $\min_{y \in S: y \neq z} |y - \mu|^2$. For every $y \neq z$,

$$\delta^2 \leq |y - z|^2 \leq |y - \mu|^2 + |\mu - z|^2 \leq |y - \mu|^2 + \sigma^2.$$

Because $\sigma^2 \leq \delta^2/2$, we have

$$\mathop{\mathbb{E}}_{T}\Big[\mathop{\mathrm{Var}}_{x}\big[\mathop{\mathbb{E}}_{U}[f(x + T \wedge U)]\big]\Big] \leq 4\left(\frac{\mathrm{Var}[f]}{\delta^2 - \sigma^2}\right)^{\frac{2}{1+\rho}} \leq 4\left(\frac{2\,\mathrm{Var}[f]}{\delta^2}\right)^{\frac{2}{1+\rho}}. \qquad \square$$

Remark 41. *The dependence on δ is necessary. Consider a function f with support $\{0, \varepsilon\}$. Then $f = \varepsilon g$, where g has support $\{0, 1\}$. We have $\mathrm{Var}[f] = \varepsilon^2 \mathrm{Var}[g]$. Applying noise to f is the same as applying noise to g, but g has no dependence on ε.*

5.2 XOR Lemma for bounded independence

We now prove a version of XOR lemma for bounded independence that is similar to the one in [GMR$^+$12], which proves the lemma for small-bias distributions.

Lemma 42. *Let $f_1, \ldots, f_k\colon \{0,1\}^m \to [0,1]$ be k functions on disjoint inputs. Let $H\colon [0,1]^k \to [0,1]$ be a multilinear function in its input. If each f_i is fooled by any d_i-wise independent distribution with error ε, then the function $h\colon \{0,1\}^m \to [0,1]$ defined by $h(x) := H(f_1(x), f_2(x), \ldots, f_k(x))$ is fooled by any $(\sum_{i \leq k} d_i)$-wise independent distribution with error $16^k \varepsilon$.*

We will use the following dual equivalence between bounded independence and sandwiching polynomials that was introduced by Bazzi [Baz07].

Fact 43 ([Baz07]). *A function $f\colon \{0,1\}^m \to [0,1]$ is fooled by every d-wise independent distribution if and only if there exist two multivariate polynomials p_ℓ and p_u of degree d such that*

1. *For every $x \in \{0,1\}^m$, we have $p_\ell(x) \leq f(x) \leq p_u(x)$, and*
2. *$\mathbb{E}[p_u(U) - f(U)] \leq \varepsilon$ and $\mathbb{E}[f(U) - p_\ell(U)] \leq \varepsilon$.*

Proof of Lemma 42. By Fact 43, for each $i \in \{1, \ldots, k\}$, there exist two degree-d_i polynomials f_i^u and f_i^ℓ for f_i which satisfy the conditions in Fact 43. Hence, we have

$$f_i^u(x) \geq f_i(x) \geq 0 \quad \text{and} \quad 1 - f_i^\ell(x) \geq 1 - f_i(x) \geq 0.$$

For every $\alpha \in \{0,1\}^k$, define

$$M_\alpha^u(x) := \prod_{i:\alpha_i = 1} f_i^u(x) \prod_{j:\alpha_j = 0} \big(1 - f_j^\ell(x)\big) \quad \text{and} \quad M_\alpha(x) := \prod_{i:\alpha_i = 1} f_i(x) \prod_{j:\alpha_j = 0} \big(1 - f_j(x)\big).$$

Clearly, $M_\alpha^u(x) \geq M_\alpha(x)$, and $M_\alpha^u(x)$ has degree $\sum_{i \leq k} d_i$. We claim that for every $\alpha \in \{0,1\}^k$,

$$\mathbb{E}[M_\alpha^u(x) - M_\alpha(x)] \leq 2^k \varepsilon.$$

Fix a string $\alpha \in \{0,1\}^k$. Define the hybrids $M_0 = M^u_\alpha(x), M_1, \dots, M_k = M_\alpha(x)$, where

$$M_i(x) := M_i^{(1)}(x) \cdot M_i^{(2)}(x),$$

where

$$M_i^{(1)}(x) := \prod_{j \leq i, \alpha_j = 1} f_j(x) \prod_{j \leq i : \alpha_j = 0} \bigl(1 - f_j(x)\bigr),$$

and

$$M_i^{(2)}(x) := \prod_{j > i : \alpha_j = 1} f^u_j(x) \prod_{j > i : \alpha_j = 0} \bigl(1 - f^\ell_j(x)\bigr).$$

Note that

$$\mathbb{E}[M_i^{(2)}(x)] = \prod_{j > i : \alpha_j = 1} \mathbb{E}\bigl[f^u_j(x)\bigr] \prod_{j > i : \alpha_j = 0} \mathbb{E}\bigl[\bigl(1 - f^\ell_j(x)\bigr)\bigr] \leq (1+\varepsilon)^{k-i},$$

and $M_i^{(1)}(x) \leq 1$. So, if $\alpha_i = 1$, then

$$\mathbb{E}\bigl[M_{i-1}(x) - M_i(x)\bigr] = \mathbb{E}\Bigl[\bigl(f^u_i(x) - f_i(x)\bigr) \cdot M_{i-1}^{(1)}(x) \cdot M_i^{(2)}(x)\Bigr] \leq \varepsilon \cdot (1+\varepsilon)^{k-i}.$$

Likewise, if $\alpha_i = 0$, we have

$$\mathbb{E}[M_{i-1}(x) - M_i(x)] = \mathbb{E}\Bigl[\bigl((1 - f^\ell_i(x)) - (1 - f_i(x))\bigr) \cdot M_{i-1}^{(1)}(x) \cdot M_i^{(2)}(x)\Bigr] \leq \varepsilon \cdot (1+\varepsilon)^{k-i}.$$

Hence,

$$\mathbb{E}[M^u_\alpha(x) - M_\alpha(x)] \leq \sum_{1 \leq i \leq k} \mathbb{E}[M_{i-1}(x) - M_i(x)] \leq \varepsilon \sum_{0 \leq i \leq k-1} (1+\varepsilon)^i \leq 2^k \varepsilon.$$

Now we define $M^\ell_\alpha(x) := 1 - \sum_{\beta : \beta \neq \alpha} M^u_\beta(x)$. Note that $M^\ell_\alpha(x)$ also has degree $\sum_{i \leq k} d_i$. Since

$$\sum_{\alpha \in \{0,1\}^k} M_\alpha(x) = \prod_{i \leq k} \bigl(f_i(x) + (1 - f_i(x))\bigr) = 1,$$

we have

$$M^\ell_\alpha(x) = 1 - \sum_{\beta : \beta \neq \alpha} M^u_\beta(x) \leq 1 - \sum_{\beta : \beta \neq \alpha} M_\beta(x) = M_\alpha(x).$$

Hence,

$$\mathbb{E}[M_\alpha(x) - M^\ell_\alpha(x)] = \sum_{\beta : \beta \neq \alpha} \bigl(M^u_\beta(x) - M_\beta(x)\bigr) \leq \sum_{\beta : \beta \neq \alpha} 2^k \varepsilon \leq 2^k 2^k \varepsilon = 4^k \varepsilon.$$

As H is multilinear, we can write H as

$$H(y_1, \dots, y_k) = \sum_{\alpha \in \{0,1\}^k} H(\alpha) \prod_{i : \alpha_i = 1} y_i \prod_{i : \alpha_i = 0} (1 - y_i),$$

where $H(\alpha) \in [0,1]$ for every α. So

$$h(x) = \sum_{\alpha \in \{0,1\}^k} H(\alpha) \prod_{i:\alpha_i = 1} f_i(x) \prod_{i:\alpha_i=0} (1 - f_i(x)) = \sum_{\alpha \in \{0,1\}^k} H(\alpha) M_\alpha(x).$$

Now if we define

$$h^u(x) := \sum_{\alpha \in \{0,1\}^k} H(\alpha) M^u_\alpha(x) \quad \text{and} \quad h^\ell(x) := \sum_{\alpha \in \{0,1\}^k} H(\alpha) M^\ell_\alpha(x).$$

Clearly h^u and h^ℓ both have degree $\sum_{i \le k} d_i$. We also have $h^u(x) \ge h(x) \ge h^\ell(x)$, and

$$\mathbb{E}[h^u(x) - h^\ell(x)] \le \sum_{\alpha \in \{0,1\}^k} H(\alpha)\, \mathbb{E}[M^u_\alpha(x) - M^\ell_\alpha(x)] \le \sum_{\alpha \in \{0,1\}^k} (2^k + 4^k)\varepsilon \le 16^k \varepsilon.$$

Therefore, since h^u and h^ℓ are two polynomials that satisfy the conditions in Fact 43, the lemma follows. □

5.3 Proof of Theorem 10

Armed with Lemma 40 and Lemma 42, we are now ready to prove Theorem 10. We first need the following useful fact to handle the case when S is the M-th roots of unity.

Fact 44. *Let X and Y be two random variables on $\{0,1\}^m$. Suppose for every product test $g\colon \{0,1\}^m \to S$, where S is the set of all M-th roots of unity, we have $\big|\mathbb{E}[g(X)] - \mathbb{E}[g(Y)]\big| \le \varepsilon$. Then for every product test $g\colon \{0,1\}^m \to S$ and every $z \in S$, we have $\big|\Pr[g(X) = z] - \Pr[g(Y) = z]\big| \le \varepsilon$.*

Proof. Note that for every integer j, the function g^j is also a product test with the same range. So for every $j, k \in \{0, \dots, m\}$,

$$\big|\mathbb{E}[(\omega^{-k} g(X))^j] - \mathbb{E}[(\omega^{-k} g(Y))^j]\big| \le \big|\omega^{-kj}\big| \cdot \big|\mathbb{E}[g(X)^j] - \mathbb{E}[g(Y)^j]\big| \le \varepsilon.$$

Using the identity

$$\frac{1}{M} \sum_{i=0}^{M-1} \omega^{(i-k)j} = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{otherwise,} \end{cases}$$

we have for every $k \in \{0, \dots, m-1\}$,

$$\begin{aligned} \big|\Pr[g(D + T \wedge U) = \omega^k] - \Pr[g(U) = \omega^k]\big| &\le \frac{1}{M} \sum_{i=0}^{M} \Big|\mathbb{E}\big[(\omega^{-k} g(D + T \wedge U))^j\big] - \mathbb{E}\big[(\omega^{-k} g(U))^j\big]\Big| \\ &\le \varepsilon. \end{aligned}$$

□

Proof of Theorem 10. Write $f = \prod_{i=1}^{k} f_i$, where $f_i \colon \{0,1\}^{I_i} \to \mathbb{C}_{\leq 1}$. Let σ denote $(\sum_{i\leq k} \mathrm{Var}[f_i])^{1/2}$. We will consider two cases: $\sigma^2 \geq dn^{0.1}$ and $\sigma^2 \leq dn^{0.1}$.

If $\sigma^2 \geq dn^{0.1}$, then the expectation of f under the uniform distribution is small. Specifically, we have

$$\Big|\prod_{i\leq k} \mathop{\mathbb{E}}_{U}[f_i(U)]\Big| = \prod_{i\leq k}(1-\mathrm{Var}[f_i])^{1/2} \leq e^{-\frac{1}{2}\sigma^2} \leq 2^{-\Omega(dn^{0.1})} \leq n^{-\Omega(d)}. \tag{14}$$

Thus, it suffices to show that its expectation under $D + T \wedge U$ is at most $n^{-\Omega(d)}$. We now use Claim 13 to show that

$$\Big|\mathop{\mathbb{E}}_{D,T,U}\Big[\prod_{i=1}^{k} f_i(D + T\wedge U)\Big]\Big| \leq n^{-\Omega(d)}.$$

For each $t, x \in \{0,1\}^m$, and each $i \in \{1, 2, \ldots, k\}$, let $\sigma^2_{t,x,i}$ denote $\mathrm{Var}_{U'}[f_i(x + t \wedge U')]$. Let T' be the uniform distribution over $\{0,1\}^m$. By Claim 13 with $\eta = 1/2$, we have $\mathbb{E}_{T',U}[\sigma^2_{T',U,i}] \geq \mathrm{Var}[f_i]/2$. So by linearity of expectation,

$$\mathop{\mathbb{E}}_{T',U}\Big[\sum_{i\leq k}\sigma^2_{T',U,i}\Big] \geq \sigma^2/2 \geq dn^{0.1}/2.$$

Since T and D are both dn-wise independent, the random variables $\sigma^2_{T,D,1}, \ldots, \sigma^2_{T,D,k}$ are $(0, dn)$-close to $\sigma^2_{T',U,1}, \ldots, \sigma^2_{T',U,k}$. Let $\mu = \mathbb{E}_{T',U}\big[\sum_{i\leq k}\sigma^2_{T',U,i}\big] \geq dn^{0.1}/2$. By Lemma 48,

$$\Pr_{T,D}\Big[\sum_{i\leq k}\sigma^2_{T,D,i} \leq \mu/2\Big] \leq 2^d\left(\frac{\sqrt{\mu d} + d}{\mu/2}\right)^d = n^{-\Omega(d)}.$$

Hence, except with probability $n^{-\Omega(d)}$ over $t \in T$ and $x \in D$, we have

$$\sum_{i\leq k}\sigma^2_{t,x,i} = \sum_{i\leq k}\mathop{\mathrm{Var}}_{U'}[f_i(x + t\wedge U')] \geq dn^{0.1}/4.$$

By a similar calculation to (14), for every such t and x,

$$\begin{aligned}
\Big|\prod_{i\leq k}\mathop{\mathbb{E}}_{U}[f_i(x + t\wedge U)]\Big| &\leq \prod_{i\leq k}\Big|\mathop{\mathbb{E}}_{U}[f_i(x + t\wedge U)]\Big| \\
&= \prod_{i\leq k}(1-\sigma^2_{t,x,i})^{1/2} \\
&\leq e^{-\frac{1}{2}\sum_{i\leq k}\sigma^2_{t,x,i}} \leq 2^{-\Omega(dn^{0.1})} \leq n^{-\Omega(d)}.
\end{aligned}$$

In addition, we always have $|f| \leq 1$. Hence,

$$\Big|\mathop{\mathbb{E}}_{D,T,U}\Big[\prod_{i\leq k} f_i(D + T\wedge U)\Big]\Big| \leq \mathop{\mathbb{E}}_{D,T}\Big[\Big|\prod_{i\leq k}\mathop{\mathbb{E}}_{U}[f_i(D + T\wedge U)]\Big|\Big] \leq n^{-\Omega(d)}.$$

Suppose $\sigma^2 \le dn^{0.1}$. Let $\sigma_1^2 \ge \sigma_2^2 \ge \cdots \ge \sigma_k^2$ be the variances of $f_1, f_2, \ldots, f_k$ respectively. Let $k' = dn^{0.7}$. We have $\sigma_{k'}^2 \le dn^{0.1}/k' = n^{-0.6}$; for otherwise $\sigma^2 \ge \sum_{i=1}^{k'} \sigma_i^2 \ge k'\sigma_{k'}^2 > dn^{0.1}$, a contradiction. Let T' be the uniform distribution over $\{0,1\}^m$. Let $\tilde{\sigma}_i^2$ denote

$$\operatorname*{Var}_{T',U}\Bigl[\mathop{\mathbb{E}}_{U'}[f(U + T' \wedge U')]\Bigr].$$

We now show that $\tilde{\sigma}_i^2 \le O(\sigma_i^2)^{4/3}$. For every $i \in \{1, \ldots, k\}$, define $g_i\colon \{0,1\}^m \to \mathbb{C}_{\le 1}$ to be $g_i(x) = (f_i(x) - \mathbb{E}[f_i])/2$ so that $\mathbb{E}[g_i] = 0$ and $\operatorname{Var}[g_i] = \operatorname{Var}[f_i]/4$. We apply Lemma 40 with $\rho = 1/2$. Notice that since M is fixed, we have $|g(x) - g(x')| = \Omega(1)$ whenever $g(x) \ne g(x')$. Hence,

$$\begin{aligned}
\tilde{\sigma}_i^2 &= \operatorname*{Var}_{T',U}\Bigl[\mathop{\mathbb{E}}_{U'}[g_i(U + T' \wedge U')]\Bigr] \\
&= \mathop{\mathbb{E}}_{T'}\Bigl[\operatorname*{Var}_{U}\bigl[\mathop{\mathbb{E}}_{U'}[g_i(U + T' \wedge U')]\bigr]\Bigr] \\
&= O(\sigma_i^2)^{4/3}.
\end{aligned}$$

It follows that

$$\sum_{i>k'} \tilde{\sigma}_i^2 = O\Bigl(\sum_{i>k'} (\sigma_i^2)^{4/3}\Bigr) \le O\bigl((\sigma_{k'}^2)^{1/3}\bigr) \sum_{i>k'} \sigma_i^2 \le O(n^{-0.2}) \cdot dn^{0.1} = dn^{-\Omega(1)}.$$

Now, if we let $F_2 := \prod_{i>k'} f_i$, then by Lemma 12,

$$\Bigl|\mathop{\mathbb{E}}_{D,T,U}[F_2(D + T \wedge U)] - \mathop{\mathbb{E}}_{U}[F_2(U)]\Bigr| \le n^{-\Omega(d)}. \tag{15}$$

On the other hand, if we define F_1 to be $\prod_{i=1}^{k'} f_i$, then it follows from Theorem 7 that

$$\Bigl|\mathop{\mathbb{E}}_{D,T,U}[F_1(D + T \wedge U)] - \mathop{\mathbb{E}}_{U}[F_1(U)]\Bigr| \le k' 2^{-\Omega(d^2 n/k')} = 2^{-\Omega(dn^{0.3})}. \tag{16}$$

We now combine (15) and (16) using Lemma 42. To begin, define $g_1(x) := \mathbb{E}_{T,U}[F_1(x + T \wedge U)]$ and $g_2(x) := \mathbb{E}_{T,U}[F_2(x + T \wedge U)]$.

If $S = [0,1]$, then the theorem follows immediately by applying Lemma 42 to g_1 and g_2. However, if S is the set of M-th roots of unity, then we cannot apply Lemma 42 directly because it only applies to functions with range $[0,1]$. Nevertheless we can use Fact 44 to reduce from S to $[0,1]$.

We now reduce S to $[0,1]$ and apply Lemma 42. For every $z \in S$, we define the point function $\mathbb{1}_z\colon S \to \{0,1\}$ by $\mathbb{1}_z(x) = 1$ if and only if $x = z$. Then for every random variable Z on S,

$$\mathbb{E}[Z] = \sum_{z \in S} z \Pr[Z = z] = \sum_{z \in S} z\, \mathbb{E}[\mathbb{1}_z(Z)].$$

Hence,

$$
\begin{aligned}
\mathbb{E}[g_1(X)g_2(X)] &= \sum_{z\in S} z\,\mathbb{E}\big[\mathbb{1}_z\big(g_1(X)g_2(X)\big)\big] \\
&= \sum_{z\in S} z\,\mathbb{E}\Big[\sum_{u,v\in S:uv=z} \mathbb{1}_u\big(g_1(X)\big)\mathbb{1}_v\big(g_2(X)\big)\Big] \\
&= \sum_{z\in S} z \sum_{u,v\in S:uv=z} \mathbb{E}\Big[\mathbb{1}_u\big(g_1(X)\big)\mathbb{1}_v\big(g_2(X)\big)\Big].
\end{aligned}
$$

Hence, by Fact 44, for every $u,v\in S$, the functions $\mathbb{1}_u\circ g_1$ and $\mathbb{1}_v\circ g_2$ are fooled by d-wise independence with error $n^{-\Omega(d)}$. So by Lemma 42, $(\mathbb{1}_u\circ f)(\mathbb{1}_v\circ g)$ are fooled by $2d$-wise independence with error $n^{-\Omega(d)}$. Hence,

$$
\begin{aligned}
&\big|\mathbb{E}[f(D+T\wedge U)]-\mathbb{E}[f(U)]\big| \\
&= \big|\mathbb{E}[(g_1g_2)(D)]-\mathbb{E}[(g_1g_2)(U)]\big| \\
&\le \sum_{z\in S}|z| \sum_{u,v\in S:uv=z} \Big|\mathbb{E}\big[(\mathbb{1}_u\circ g_1)(\mathbb{1}_v\circ g_2)\big(D\big)\big]-\mathbb{E}\big[(\mathbb{1}_u\circ g_1)(\mathbb{1}_v\circ g_2)\big(U\big)\big]\Big| \\
&\le M^2\cdot n^{-\Omega(d)} = n^{-\Omega(d)}
\end{aligned}
$$

because M is fixed, proving the theorem. □

Comparison on the amount of independence and noise with other works. We end this section with a comparison of the amount of independence and noise rate used in our work with others which also use the Ajtai–Wigderson framework.

[GMR+12] implicitly show that $\tilde{O}(\log(m/\varepsilon))$-wise independence plus noise with constant-rate fools read-once CNFs with error ε. For read-once polynomials, monomials of degree $\Omega(\log(m/\varepsilon))$ do not change the expectation by much, and so we can replace n in Theorem 8 with $\log(m/\varepsilon)$, showing that $O(\log(m/\varepsilon)\log(1/\varepsilon))$-wise independence plus constant-rate noise fools read-once polynomials with error ε. We can improve the amount of independence to $\tilde{O}(\log(m/\varepsilon))$ like [GMR+12], by combining its proof with Claim 14. However, we omit the proof here as we do not know how to exploit this improvement to obtain a better seed length for our PRG. For a simple concrete example, consider a degree-2 polynomial. Assuming in the ideal scenario that each recursion uses a seed of length $O(\log(1/\varepsilon)) = O(\log m)$. Since our progress measure is its degree, as discussed before, we need to recurse $O(\log m)$ steps to halve the degree. So in total the seed length is $O(\log^2 m)$. We note that same problem already appears when the polynomial has degree $O(\log\log m)$, which is too large to apply the PRG for low-degree polynomials [Vio09] if one wants seed length $\tilde{O}(\log m)$.

We point out that the noise rate in [GMR+12] and this work is subtly different from the ones used in the results for unordered branching programs [RSV13, SVW14, CHRT18, CHHL18].

The latter results rely on bounding the Fourier spectrum of branching programs, which implicitly show that $O(\log m)$-wise independence plus noise with rates $1-1/\Omega(\log n)^w$ and $1-$

$1/\Omega(w^2)$ fool width-w general and regular branching programs with error $n^{-\Omega(1)}$, respectively. Noise with such rates alone are known to be sufficient to fool these models with constant error [Ste13, BRRY10]. In fact, Cohen, Ganor and Raz [CGR14] showed that a weaker notion of noise with a similar rate suffices.

On the other hand, without bounded independence, by considering the Tribes function, we can see that fooling CNFs, read-once polynomials and product tests all requires $1 - 1/\Omega(\log n)$-rate noise. While these work use slightly greater amount of independence than the latter, a smaller noise rate allows fewer recursions in the construction of PRGs, and therefore gives PRGs with shorter seeds for these models.

6 Small-bias plus noise fools degree-2 polynomials

In this section we show that small-bias distributions plus noise fool non-read-once $\mathbb{F}_2$-polynomials of degree 2. We first state a structural theorem about degree-2 polynomials over $\mathbb{F}_2$ which will be used in our proof.

Theorem 45 (Theorem 6.30 in [LN97])**.** *For every $\mathbb{F}_2$-polynomial $p\colon \{0,1\}^m \to \{0,1\}$ of degree 2, there exists an invertible matrix $A \in \mathbb{F}_2^{m\times m}$, an integer $k \le \lfloor m/2 \rfloor$, and a subset $L \subseteq [m]$ such that $p(Ax) := \sum_{i=1}^k x_{2i-1}x_{2i} + \sum_{i\in L} x_i$.*

Proof of Claim 11. Let p be a degree-2 polynomial. It suffices to fool $q(x) := (-1)^{p(x)}$. By Theorem 45, there exists an invertible matrix such that $q(Ax) = r(x)\cdot\chi_L(x)$, where $r(x) := (-1)^{\sum_{i=1}^k x_{2i-1}x_{2i}}$, and $\chi_L(x) = (-1)^{\sum_{i\in L} x_i}$. By writing $r(x)$ in its Fourier expansion, $q(x)$ has the Fourier expansion

$$q(x) = \Big(\sum_{S\subseteq[k]} \hat{r}_S \chi_S(x)\Big)\chi_L(x),$$

where $|\hat{r}_S| = 2^{-k/2}$. Note that L is a subset of $[m]$. Viewing the sets S and L as vectors in $\{0,1\}^m$, we have

$$\begin{aligned}\big|\mathbb{E}[q(D+T\wedge U)] - \mathbb{E}[q(U)]\big| &\le \sum_{\emptyset\ne S\subseteq[k]} 2^{-k/2}\big|\mathbb{E}[\chi_{S+L}(A^{-1}(D))]\big|\cdot\big|\mathbb{E}[\chi_{S+L}(A^{-1}(T\wedge U))]\big| \\ &\le 2^{-k/2}\delta \sum_{\emptyset\ne S\subseteq[k]} \big|\mathbb{E}[\chi_{S+L}(A^{-1}(T\wedge U))]\big| \\ &= 2^{-k/2}\delta \sum_{\emptyset\ne S\subseteq[k]} \big|\mathbb{E}[\chi_{A(S+L)}(T\wedge U)]\big| \\ &= 2^{-k/2}\delta \sum_{\emptyset\ne S\subseteq[k]} (1/3)^{|A(S+L)|},\end{aligned}$$

where the second inequality follows because small-bias distributions are closed under linear transformations. We now bound above the summation. We claim that

$$\sum_{S\subseteq[k]} (1/3)^{|A(S+L)|} \le \sum_{S\subseteq[k]} (1/3)^{|S|} = (4/3)^k.$$

The equality is clear. To see the inequality, notice that since $S \subseteq [k]$, when viewed as a vector in $\{0,1\}^m$ its last $m-k$ positions must be 0. So we will instead think of S as a vector in $\{0,1\}^k$, and rewrite $A(S+L)$ as $A'S+AL$, where A' is the first k columns of the full rank matrix A. In particular, A' is a full rank $m \times k$ matrix. As we are only concerned with the Hamming weight of $A'S+AL$, we can permute its coordinates and rewrite A' as $[I_k|A'']^T$ for some $k \times (m-k)$ matrix A''. (Readers who are familiar with linear codes should think of the standard form of a generator matrix.) Moreover, for a lower bound on the Hamming weight, we can restrict our attention to the first k bits of $A'S+AL$. Hence, we can think of first k bits of $A'S+AL$ as S shifted by the first k bits of the fixed vector AL. Since we are summing over all S in $\{0,1\}^k$, the shift does not affect the sum, and the inequality follows. Therefore, we have

$$\left|\mathbb{E}[q(D+T \wedge U)] - \mathbb{E}[q(U)]\right| = 2^{-k/2}\delta \cdot (4/3)^k \leq (8/9)^{k/2}\delta,$$

and proving the claim. □

7 Proof of Claim 9

In this section, we more generally exhibit a distribution D that is $(d^2/10k, d)$-close to uniform. One can obtain Claim 9 by setting $d = k^{1/3}$. To simplify notation we will switch from $\{0,1\}$ to $\{-1,1\}$, and replace k with $2k$.

We define D to be the uniform distribution over strings in $\{-1,1\}^{2k}$ with equal number of -1's and 1's.

Claim 46. *D is $(10d^2/k, d)$-close to uniform for every integer d.*

Proof. We can assume $d^2 \leq k/10$, for otherwise the conclusion is trivial. Let $I \subseteq [k]$ be a subset of size d. For every $x \in \{-1,1\}^d$, we have

$$\Pr[D_I = x] = \frac{\binom{2k-d}{k-wt(x)}}{\binom{2k}{k}},$$

where $wt(x)$ is the number of -1's in x. We bound below the R.H.S. by

$$\begin{aligned}\frac{\binom{2k-d}{k-d}}{\binom{2k}{k}} &= \frac{k(k-1)\cdots(k-d+1)}{2k(2k-1)\cdots(2k-d+1)} \\ &\geq \left(\frac{k-d+1}{2k}\right)^d \\ &= 2^{-d}\left(1-\frac{d-1}{k}\right)^d \\ &\geq 2^{-d}\left(1-\frac{d(d-1)}{k}\right) \\ &\geq 2^{-d}\cdot(1-d^2/k),\end{aligned}$$

and bound it above by

$$\begin{aligned}\frac{\binom{2k-d}{k-d/2}}{\binom{2k}{k}} &= \frac{(k(k-1)\cdots(k-d/2+1))^2}{2k(2k-1)\cdots(2k-d+1)}\\ &\le \left(\frac{k}{2k-d+1}\right)^d\\ &= 2^{-d}\left(1+\frac{d-1}{2k-d+1}\right)^d\\ &\le 2^{-d}\left(1+\sum_{i=1}^{d}\left(\frac{d(d-1)}{2k-d+1}\right)^i\right)\\ &\le 2^{-d}\left(1+2\cdot\frac{d(d-1)}{2k-d+1}\right)\\ &\le 2^{-d}\cdot(1+2d^2/k).\end{aligned}$$

The third inequality is because the geometric sum has ratio $\le 1/2$ as $d^2 \le k/10$, and so is bounded by twice the first term. Hence, we have $|\Pr[D_I = x] - 2^{-d}| \le 2^{-d}\cdot 2d^2/k$ for every $x \in \{-1,1\}^d$. The claim then follows from summing the inequality over every $x \in \{-1,1\}^d$. □

We now define our product test f. For each $j \in \{1,\dots,2k\}$, define $f_j\colon \{-1,1\}^{2k} \to \mathbb{C}_{\le 1}$ to be $f_j(x) = \omega^{x_j}$, where $\omega := e^{-i/\sqrt{2k}}$. Let $f = \prod_{j\le 2k} f_j$. We now show that for every large enough k we have

$$\Big|\mathbb{E}[f(D+T\wedge U)] - \mathbb{E}[f(U)]\Big| \ge 1/10.$$

We now bound above and below the expectation of f under both distributions. We will use the fact that $1-\theta^2/2 \le \cos\theta \le 1-2\theta^2/5$ for $\theta\in[-1,1]$. First, we have

$$\mathbb{E}[f(U)] = \prod_{j\le 2k}\mathop{\mathbb{E}}_{x\sim\{-1,1\}}[\omega^x] = \prod_{j\le 2k}(\omega+\omega^{-1})/2 = \Big(\cos(1/\sqrt{2k})\Big)^{2k} \le (1-1/5k)^{2k}.$$

Next for every $j\in\{1,2,\dots,2k\}$, we have

$$\mathop{\mathbb{E}}_{T,U}[f_j(x+T\wedge U)] = \frac{3}{4}\omega^{x_j} + \frac{1}{4}\omega^{-x_j}.$$

Define $\beta\colon\{-1,1\}\to\mathbb{C}_{\le 1}$ to be $\beta(x) := \frac{3}{4}\omega^x + \frac{1}{4}\omega^{-x}$. Since D has the same number of -1's and 1's,

$$\begin{aligned}\mathop{\mathbb{E}}_{D}\Big[\prod_{j\le 2k}\beta_j(D)\Big] &= \beta(1)^k\beta(-1)^k\\ &= (10/16 + 3/16\cdot(\omega^2+\omega^{-2}))^k\\ &= (5/8 + 3/8\cdot\cos(2/\sqrt{2k}))^k\\ &\ge (5/8+3/8\cdot(1-1/k))^k\\ &= (1-3/8k)^k,\end{aligned}$$

Therefore $|\mathbb{E}[f(D+T\wedge U]-\mathbb{E}[f(U)]| \geq (1-3/8k)^k-(1-1/5k)^{2k} \geq 1/10$, for every sufficiently large k, concluding the proof.

The f_i in this proof have variance $\Theta(1/k)$. So this counterexample gives a product test with total-variance $O(1)$, and is relevant also to Lemma 12. Specifically it shows that for $n=1$ and say $d=O(1)$, the error term $(k2^n)^{O(d)}\varepsilon$ in Lemma 12 cannot be replaced with $k^c\varepsilon$ for a certain constant c. Moreover, it cannot be replaced even if any $k^{\Omega(1)}$ of the Y_i are close to the X_i (as opposed to just $O(1)$).

Acknowledgments. We thank Daniel Kane for answering some questions about [GKM15].

References

[AKS87] Miklós Ajtai, János Komlós, and Endre Szemerédi. Deterministic simulation in logspace. In *19th ACM Symp. on the Theory of Computing (STOC)*, pages 132–140, 1987.

[ASWZ96] Roy Armoni, Michael E. Saks, Avi Wigderson, and Shiyu Zhou. Discrepancy sets and pseudorandom generators for combinatorial rectangles. In *37th IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 412–421, 1996.

[AW89] Miklos Ajtai and Avi Wigderson. Deterministic simulation of probabilistic constant-depth circuits. *Advances in Computing Research - Randomness and Computation*, 5:199–223, 1989.

[Baz07] Louay Bazzi. Polylogarithmic independence can fool DNF formulas. In *48th IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 63–73, 2007.

[BPW11] Andrej Bogdanov, Periklis A. Papakonstantinou, and Andrew Wan. Pseudorandomness for read-once formulas. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 240–246, 2011.

[BRRY10] Mark Braverman, Anup Rao, Ran Raz, and Amir Yehudayoff. Pseudorandom generators for regular branching programs. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, 2010.

[BV10] Andrej Bogdanov and Emanuele Viola. Pseudorandom bits for polynomials. *SIAM J. on Computing*, 39(6):2464–2486, 2010.

[CGR14] Gil Cohen, Anat Ganor, and Ran Raz. Two sides of the coin problem. In *Workshop on Randomization and Computation (RANDOM)*, pages 618–629, 2014.

[CHHL18] Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, and Shachar Lovett. Pseudorandom generators from polarizing random walks. *Electronic Colloquium on Computational Complexity*, Technical Report TR18-015, 2018. www.eccc.uni-trier.de/.

[CHRT18] Eshan Chattopadhyay, Pooya Hatami, Omer Reingold, and Avishay Tal. Improved pseudorandomness for unordered branching programs through local monotonicity. In *ACM Symp. on the Theory of Computing (STOC)*, 2018.

[CRS00] Suresh Chari, Pankaj Rohatgi, and Aravind Srinivasan. Improved algorithms via approximations of probability distributions. *J. Comput. System Sci.*, 61(1):81–107, 2000.

[De15] Anindya De. Beyond the central limit theorem: Asymptotic expansions and pseudorandomness for combinatorial sums. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 883–902, 2015.

[EGL$^+$92] Guy Even, Oded Goldreich, Michael Luby, Noam Nisan, and Boban Velickovic. Approximations of general independent distributions. In *ACM Symp. on the Theory of Computing (STOC)*, pages 10–16, 1992.

[EGL$^+$98] Guy Even, Oded Goldreich, Michael Luby, Noam Nisan, and Boban Velickovic. Efficient approximation of product distributions. *Random Struct. Algorithms*, 13(1):1–16, 1998.

[GKM15] Parikshit Gopalan, Daniel Kane, and Raghu Meka. Pseudorandomness via the discrete fourier transform. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 903–922, 2015.

[GLS12] Dmitry Gavinsky, Shachar Lovett, and Srikanth Srinivasan. Pseudorandom generators for read-once accˆ0. In *IEEE Conf. on Computational Complexity (CCC)*, pages 287–297, 2012.

[GMR$^+$12] Parikshit Gopalan, Raghu Meka, Omer Reingold, Luca Trevisan, and Salil Vadhan. Better pseudorandom generators from milder pseudorandom restrictions. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, 2012.

[GMRZ13] Parikshit Gopalan, Raghu Meka, Omer Reingold, and David Zuckerman. Pseudorandom generators for combinatorial shapes. *SIAM J. on Computing*, 42(3):1051–1076, 2013.

[GOWZ10] Parikshit Gopalan, Ryan O'Donnell, Yi Wu, and David Zuckerman. Fooling functions of halfspaces under product distributions. In *25th IEEE Conf. on Computational Complexity (CCC)*, pages 223–234. IEEE, 2010.

[GY14] Parikshit Gopalan and Amir Yehudayoff. Inequalities and tail bounds for elementary symmetric polynomials. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:19, 2014.

[Hat14] Hamed Hatami. Lecture notes on harmonic analysis of boolean functions, 2014. Available at `http://cs.mcgill.ca/~hatami/comp760-2014/lectures.pdf`.

[HLV17] Elad Haramaty, Chin Ho Lee, and Emanuele Viola. Bounded independence plus noise fools products. In *Conf. on Computational Complexity (CCC)*, 2017.

[INW94] Russell Impagliazzo, Noam Nisan, and Avi Wigderson. Pseudorandomness for network algorithms. In *26th ACM Symp. on the Theory of Computing (STOC)*, pages 356–364, 1994.

[LN97] Rudolf Lidl and Harald Niederreiter. *Finite fields*, volume 20 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, second edition, 1997.

[Lov09] Shachar Lovett. Unconditional pseudorandom generators for low degree polynomials. *Theory of Computing*, 5(1):69–82, 2009.

[Lu02] Chi-Jen Lu. Improved pseudorandom generators for combinatorial rectangles. *Combinatorica*, 22(3):417–433, 2002.

[LVW93] Michael Luby, Boban Veličković, and Avi Wigderson. Deterministic approximate counting of depth-2 circuits. In *2nd Israeli Symposium on Theoretical Computer Science (ISTCS)*, pages 18–24, 1993.

[Nis91] Noam Nisan. Pseudorandom bits for constant depth circuits. *Combinatorica. An Journal on Combinatorics and the Theory of Computing*, 11(1):63–70, 1991.

[Nis92] Noam Nisan. Pseudorandom generators for space-bounded computation. *Combinatorica*, 12(4):449–461, 1992.

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: efficient constructions and applications. *SIAM J. on Computing*, 22(4):838–856, 1993.

[NZ96] Noam Nisan and David Zuckerman. Randomness is linear in space. *J. of Computer and System Sciences*, 52(1):43–52, February 1996.

[RSV13] Omer Reingold, Thomas Steinke, and Salil P. Vadhan. Pseudorandomness for regular branching programs via Fourier analysis. In *Workshop on Randomization and Computation (RANDOM)*, pages 655–670, 2013.

[ST18] Rocco A. Servedio and Li-Yang Tan. Improved pseudorandom generators from pseudorandom multi-switching lemmas. *CoRR*, abs/1801.03590, 2018.

[Ste04] J. Michael Steele. *The Cauchy-Schwarz master class*. MAA Problem Books Series. Mathematical Association of America, Washington, DC; Cambridge University Press, Cambridge, 2004.

[Ste13] John P. Steinberger. The distinguishability of product distributions by read-once branching programs. In *IEEE Conf. on Computational Complexity (CCC)*, pages 248–254, 2013.

[SVW14] Thomas Steinke, Salil P. Vadhan, and Andrew Wan. Pseudorandomness and fourier growth bounds for width-3 branching programs. In *Workshop on Randomization and Computation (RANDOM)*, pages 885–899, 2014.

[Tre10] Luca Trevisan. Open problems in unconditional derandomization. Presentation at China Theory Week, 2010. Slides available at `https://www.cc.gatech.edu/~mihail/trevisan2.pdf`.

[Tzu09] Yoav Tzur. Notions of weak pseudorandomness and $GF(2^n)$-polynomials. *M.Sc. thesis, Weizmann Institute of Science*, 2009.

[Vio07] Emanuele Viola. Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates. *SIAM J. on Computing*, 36(5):1387–1403, 2007.

[Vio09] Emanuele Viola. The sum of d small-bias generators fools polynomials of degree d. *Computational Complexity*, 18(2):209–217, 2009.

[Vio14] Emanuele Viola. Randomness buys depth for approximate counting. *Computational Complexity*, 23(3):479–508, 2014.

[Vio17] Emanuele Viola. Special topics in complexity theory. Lecture notes of the class taught at Northeastern University. Available at http://www.ccs.neu.edu/home/viola/classes/spepf17.html, 2017.

A Moment bounds for sum of almost d-wise independent variables

In this section we prove some moment bounds and tail bounds for sum of almost d-wise independent complex variables.

Lemma 47. *Let $Z_1, Z_2, \ldots, Z_k \in \mathbb{C}$ be independent random variables with $\mathbb{E}[Z_i] = 0$, $|Z_i| < B$. Let d be an even positive integer. Let $W_1, W_2, \ldots, W_k \in \mathbb{C}$ be random variables that are (ε, d)-close to $Z_1, \ldots, Z_k$. Then,*

$$\mathbb{E}\left[\Big|\sum_{i=1}^k W_i\Big|^d\right] \le 2^d\Big(\Big(\sum_i \mathrm{Var}[Z_i]\cdot d\Big)^{1/2} + dB\Big)^d + (2kB)^d\varepsilon.$$

Proof of Lemma 47. Note that for any random variable $W \in \mathbb{C}$ we have

$$\begin{aligned}\mathbb{E}\left[|W|^d\right] &= \mathbb{E}\left[\Big(|\Re(W)|^2 + |\Im(W)|^2\Big)^{d/2}\right] \\ &\le \mathbb{E}\left[\Big(2\max\{|\Re(W)|^2, |\Im(W)|^2\}\Big)^{d/2}\right] \\ &\le 2^{d/2}\cdot\mathbb{E}\left[|\Re(W)|^d + |\Im(W)|^d\right],\end{aligned}$$

and $\mathrm{Var}[W] = \mathrm{Var}[\Re(W)] + \mathrm{Var}[\Im(W)]$. We will first prove the lemma when W is real-valued.

Since $W_1, \ldots, W_k$ are (ε, d)-close to $Z_1, \ldots, Z_k$, and d is even, we have

$$\begin{aligned}\mathbb{E}\left[\Big|\sum_{i=1}^k W_i\Big|^d\right] &= \mathbb{E}\left[\Big(\sum_i W_i\Big)^d\right] \\ &\le \sum_{i_1,\ldots,i_d}\mathbb{E}\left[\prod_{j=1}^d Z_{i_j}\right] + k^d B^d\varepsilon,\end{aligned}$$

because there are k^d products in the sum, each product is bounded by B^d and Claim 22. We now estimate the quantity $\sum_{i_1,\ldots,i_d}\mathbb{E}\left[\prod_{j=1}^d Z_{i_j}\right]$. We have

$$\sum_{i_1,\ldots,i_d}\mathbb{E}\left[\prod_{j=1}^d Z_{i_j}\right] = \sum_{m=1}^d\sum_{|S|=m}\sum_{\substack{i_1,\ldots,i_d\in S:\\ \{i_j\}_j=S}}\mathbb{E}\left[\prod_{j=1}^d Z_{i_j}\right].$$

The expectation is zero whenever Z_{i_j} appears only once for some $i_j \in S$. So each Z_{i_j} must appear at least twice. So the expectation is 0 whenever $m > d/2$. As each Z_i is bounded by B, each product is bounded by $B^{d-2m}\prod_{j\in S}\mathbb{E}[Z_j^2] = B^{d-2m}\prod_{j\in S}\mathrm{Var}[Z_j]$. For each $S \subseteq [k]$ of size m, there are at most m^d such terms. Let σ denote $(\sum_{i=1}^k \mathrm{Var}[Z_i])^{1/2}$. Then,

$$\begin{aligned}
\sum_{i_1,\ldots,i_d} \mathbb{E}\left[\prod_{j=1}^d Z_{i_j}\right] &\le \sum_{m=1}^{d/2} B^{d-2m} m^d \sum_{|S|=m}\prod_{j\in S}\mathrm{Var}[Z_j] \\
&\le \sum_{m=1}^{d/2} B^{d-2m} m^{d-m} e^m \sigma^{2m} && \text{(Maclaurin's inequality, see Claim 23)} \\
&\le e^{d/2}\sum_{m=1}^{d/2} B^{d-2m}(d/2)^{d-m}\sigma^{2m} \\
&\le e^{d/2}(d/2)^d B^d \sum_{m=0}^{d/2}\Big(\frac{\sigma^2}{(d/2)B^2}\Big)^m \\
&\le e^{d/2}(d/2)^d B^d \cdot \left(d\Big(1+\frac{\sigma^d}{(d/2)^{d/2}B^d}\Big)\right) && \Big(\sum_{m=0}^{d-1}\alpha^m \le d(\alpha^0+\alpha^{d-1}), \forall \alpha > 0\Big) \\
&\le de^{d/2}\Big((d/2)^d B^d + (d/2)^{d/2}\sigma^d\Big) \\
&\le 2^{d/2}(dB+\sigma\sqrt{d})^d.
\end{aligned}$$

Putting everything together, we have

$$\begin{aligned}
\mathbb{E}\left[\Big|\sum_{i=1}^k W_i\Big|^d\right] &\le 2^{d/2}\left(2^{d/2}(\sigma\sqrt{d}+dB)^d + (kB)^d\varepsilon\right) \\
&\le 2^d(\sigma\sqrt{d}+dB)^d + (2kB)^d\varepsilon.
\end{aligned}$$

□

Lemma 48. *Let $X_1, X_2, \ldots, X_k \in [0,1]$ be independent random variables. Let d be an even positive integer. Let $Y_1, Y_2, \ldots, Y_k \in [0,1]$ be random variables that are (ε, d)-close to $X_1, \ldots, X_k$. Let $Y = \sum_{i\le k} Y_i$ and $\mu = \mathbb{E}[\sum_i X_i]$. Then,*

$$\Pr[|Y-\mu| \ge \delta\mu] \le 2^d\left(\frac{\sqrt{\mu d}+d}{\delta\mu}\right)^d + \left(\frac{2k}{\delta\mu}\right)^d \varepsilon.$$

In particular, if $\mu \ge 25d$ and $\delta = 1/2$, we have $\Pr[|Y-\mu| \ge \mu/2] \le 2^{-\Omega(d)} + k^d\varepsilon$.

Proof. Let $X_i' = X_i - \mathbb{E}[X_i]$, $Y_i' = Y_i - \mathbb{E}[X_i]$ and $Y' = \sum_i Y_i'$. Note that $X_i' \in [-1,1]$ and $\mathbb{E}[X_i'] = 0$. Since $X_i \in [0,1]$, we have

$$\mathbb{E}[X_i] \ge \mathbb{E}[X_i^2] \ge \mathrm{Var}[X_i] = \mathrm{Var}[X_i - \mathbb{E}[X_i]] = \mathrm{Var}[X_i'].$$

By Lemma 47 and Markov's inequality,

$$\begin{aligned}\Pr[|Y-\mu| \geq \delta\mu] &= \Pr[|Y'|^d \geq (\delta\mu)^d] \\ &\leq 2^d \left(\frac{(\sum_i \mathrm{Var}[X_i'] \cdot d)^{1/2} + d}{\delta\mu}\right)^d + \left(\frac{2k}{\delta\mu}\right)^d \varepsilon \\ &\leq 2^d \left(\frac{\sqrt{\mu d} + d}{\delta\mu}\right)^d + \left(\frac{2k}{\delta\mu}\right)^d \varepsilon,\end{aligned}$$

where in the last inequality we used $\mu \geq \sum_i \mathrm{Var}[X_i']$. □